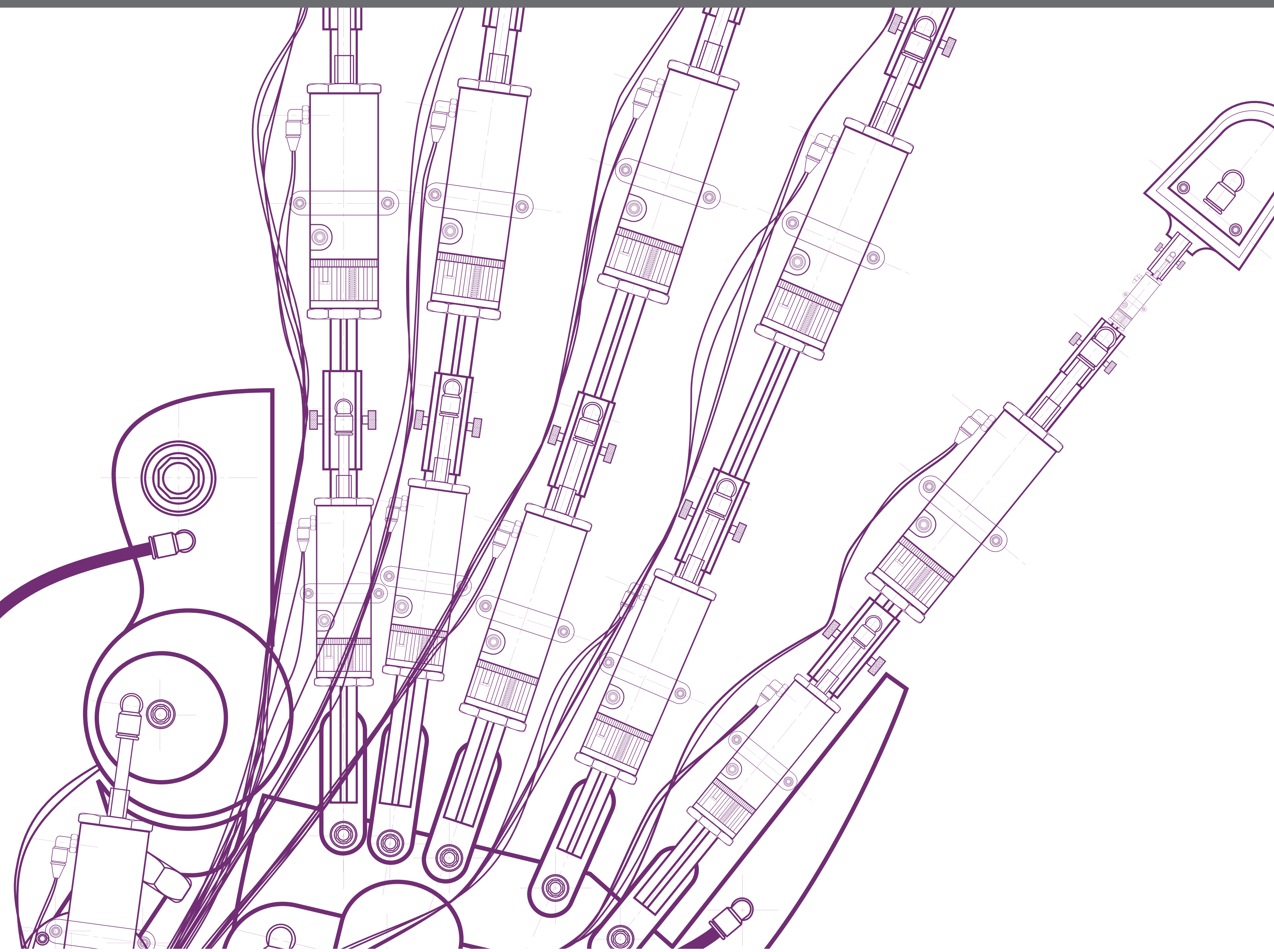
RISING STARS IN HUMAN-ROBOT INTERACTION

EDITED BY: Bilge Mutlu, Ginevra Castellano, James Everett Young, Séverin Lemaignan and Adriana Tapus

PUBLISHED IN: Frontiers in Robotics and AI



frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ISSN 1664-8714
ISBN 978-2-88976-757-1
DOI 10.3389/978-2-88976-757-1

RISING STARS IN HUMAN-ROBOT INTERACTION

Topic Editors:
Bilge Mutlu, University of Wisconsin-Madison, United States
Ginevra Castellano, Uppsala University, Sweden
James Everett Young, University of Manitoba, Canada
Séverin Lemaignan, Pal Robotics S.L. Barcelona, Spain
Adriana Tapus, École Nationale Supérieure de Techniques Avancées Palaiseau, France

Citation: Mutlu, B., Castellano, G., Young, J. E., Lemaignan, S., Tapus, A., eds. (2022). Rising Stars in Human-Robot Interaction. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-757-1

Table of Contents

frontiers in Robotics and AI

REVIEW
published: 15 July 2021
doi: 10.3389/frobt.2021.701938

Robot-Guided Evacuation as a Paradigm for Human-Robot Interaction Research

*Alan R. Wagner**

Robot Ethics and Aerial Vehicles Laboratory, Department of Aerospace Engineering, The Pennsylvania State University, University Park, PA, United States

This paper conceptualizes the problem of emergency evacuation as a paradigm for investigating human-robot interaction. We argue that emergency evacuation offers unique and important perspectives on human-robot interaction while also demanding close attention to the ethical ramifications of the technologies developed. We present a series of approaches for developing emergency evacuation robots and detail several essential design considerations. This paper concludes with a discussion of the ethical implications of emergency evacuation robots and a roadmap for their development, implementation, and evaluation.

Keywords: evacuation, human-robot interaction, robot ethics, emergency robotics, human-robot trust

OPEN ACCESS

Edited by:
James Everett Young, University of Manitoba, Canada

Reviewed by:
Jimin Rhim, McGill University, Canada
Astrid Marieke Rosenthal-Von Der Pütten, RWTH Aachen University, Germany

Correspondence:
Alan R. Wagner
alan.r.wagner@psu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *28 April 2021*
Accepted: *30 June 2021*
Published: *15 July 2021*

Citation:
Wagner AR (2021) Robot-Guided Evacuation as a Paradigm for Human-Robot Interaction Research. Front. Robot. AI 8:701938. doi: 10.3389/frobt.2021.701938

INTRODUCTION

The field of Human-Robot Interaction (HRI) tends to focus on service, education, entertainment, and healthcare applications (Bartneck, et al., 2020). HRI applications in these areas lend themselves to laboratory development and eventual experimentation in real-world settings. Moreover, for the most part, the human that is the focus of the service, educational experience, being entertained or whose health is being attended to, is typically relaxed, affable, attentive to the robot, and contemplative of the robot's performance on its assigned tasks.

The psychological disposition and resulting behavior of people using a robot for these applications stands in stark contrast to the use of robots for search and rescue or emergency evacuation applications. During search and rescue or emergency evacuation people tend to be emotional, tense, confused, inattentive, pliable, and reactive to the robot without consideration of its performance. In other words, rescue and emergency evacuation situations tend to put people in a different state of mind than traditional HRI application areas. Although, on the surface it may appear as if a dichotomy exists between applications in these two areas, in reality people's behavior can differ from day-to-day. While being served, taught, entertained, or treated, occasionally people will be emotional, tense, and reactive. It therefore behooves the HRI community to consider and explore both sides of the human state of mind in order to develop robots that might be capable of prolonged interaction with people and responsive to their daily psychological states.

With this in mind, our research examines the development and use of mobile robots as guides leading human evacuees to safety. We focus on the evacuation of buildings that contain large numbers of people, high-rise residential complexes, schools, and shopping malls, for example, because we believe that emergency guidance robots placed in these buildings could save a significant number of lives. At least with respect to high-rise residential complexes, the global number of these buildings is increasing (CTBUH, 2018) and the evacuation of these buildings is a complex and time-consuming process (Gershon et al., 2007). For example, after the 1993 World Trade Center bombing

FIGURE 1 | Confusing exit signs.

people could be found at their desk 6 h after the beginning of the evacuation (Fahy RF, 1995). Many high-rise buildings are not designed for rapid emergency evacuation (Meacham, 1999). Moreover, rescue of occupants by first responders is dangerous and takes a long time after the beginning of the emergency (Gershon et al., 2012).

In its broadest formulation, robot-guided emergency evacuation tasks the robot with leading individuals or groups of people away from danger to a safe location. This broad formulation can, however, be delimited in a variety of different ways to make the problem more tractable without necessarily compromising real-world applicability. For example, if the robot is tasked with evacuating residents from an apartment building, then the robot can be provided with a priori information about the location of exits, stairwells, and potential evacuees. Moreover, many evacuation environments have ample infrastructure, such as WIFI, to make the task easier. These and other reasonably grounded assumptions simplify the robot-guided emergency evacuation task. Traditional, non-robotic approaches to emergency evacuation include the use of exit signs, flooring lights, and broadcast announcements. These approaches are reasonably effective as a means of communicating a pathway to an exit. But traditional approaches to emergency evacuation are static and may not be well informed about the emergency. For example, during the 2001 World Trade Center bombing announcements over the public address systems told evacuees to return to their desks (Averill et al., 2013). Exit signs can be confusing (**Figure 1**) and some exits may be blocked or overcrowded. Robot emergency evacuation guides may therefore be able to adapt to the emergency in real time to save lives.

As with many HRI problems, robot-evacuee interaction is dictated by the context, the evacuee(s), and the robot. Contextual factors include the cause and type of emergency, the location of the evacuation, and the ease of exiting. Evacuation from a school, for example, is physically less demanding than evacuation from a high-rise building because a large number of stairs must traversed in order exit a high-rise building. The type of evacuee includes factors such as the person's age, disabilities, or psychological state. For instance, evacuating children may require different methods and styles of communication than evacuating adults. Health and especially health limitations can affect one's ability to evacuate and understand the robot's directions. Finally, factors related to the robot include how the robot should be designed, the modalities it uses to communicate with evacuees, and how it uses its communicative behaviors and mobility to successfully and quickly evacuate people.

The remainder of this paper begins by presenting a rationale for robot-guided evacuation. Next we describe different approaches to robot-guided evacuation, including a discussion of how to formulate the problem of robot-guided evacuation. We then discuss principles for robot design followed by an examination of the ethical implications of robot-assisted evacuation. We conclude by presenting a roadmap for this application domain.

WHY EVACUATION?

There are a number of reasons that robot-guided emergency evacuation is a valuable human-robot interaction problem. First, and perhaps most importantly, robots might one day be developed to serve as an instantaneous first responders immediately reacting to an emergency by contacting the authorities and guiding people to safety. During an emergency, people are often initially confused (Proulx, 2003). Information and leadership, even in the form of a robot, may be important for initiating the evacuation process (Bryan, 2002). Ideally, these robots will save lives by reducing the time required for evacuation, increasing the number of people evacuated, reducing crowding at exits, and providing timely information about the emergency to evacuees.

Evacuation robots might not only protect the lives of evacuees but also reduce the risks for first responders. By providing information about the emergency to first responders the robot might be able to alleviate some of the risks to first responders. For instance, simply providing camera images or streaming video could help first responders gauge the nature of the situation. Moreover, robots might be developed that could be remotely controlled, thereby allowing first responders to intentionally gather information about the emergency or how evacuees are responding to the emergency. One can imagine an advanced 911 operator that responds to emergencies in apartment buildings or schools by taking control of an onsite robot to provide additional up-to-date information for police officers and fire fighters.

Robot-guided evacuation also allows HRI researchers to investigate how humans in a highly aroused and potentially emotional state of mind interact with a robot. The vast majority of HRI research focuses on applications developed for staid, controlled environments such as one's home or the classroom (Kidd and Breazeal, 2007; Park et al., 2017; Zachiotis et al., 2018). Very little HRI research has examined situations in which the human or humans are under pressure or threat of physical harm. People act differently during an emergency (Klein et al., 1986; Jansen et al., 1995). Interacting with a person that is fleeing from some threat is, in many ways, fundamentally different from interacting with a person in a laboratory environment. Fight-or-flight responses can be debilitating and impair judgment. Evacuee-robot interaction may therefore demand a unique perspective on how a robot should behave. An evacuation robot may need to adjust its behavior based on the person's reaction and emotional state; it may need to be authoritative and interact with a commanding presence in order to convince people when and how to leave (Kuligowski, 2008; Robinette et al., 2012). The dynamic nature of the evacuee-robot relationship presents challenges as well as opportunities for important and novel research.

For many HRI applications, ecological validity is simply assumed (Dragan et al., 2013; Bartneck, et al., 2020; Chen et al., 2018). It may be expected that HRI research performed in a well-controlled laboratory experiment will extrapolate to more realistic settings. For some applications, such as service robots, such assumptions may be warranted (King et al., 2010; Mast, et al., 2015). For applications such as emergency evacuation, on the other hand, researchers cannot assume that results and data gathered from laboratory experiments will inform how real people react to a real emergency. Because externally invalid simplifications could eventually increase the risks to evacuees, we argue that research in this area must include real-world experiments with real robots operating as they would during a real emergency. Although challenging, these experiments serve to moor simulation experiments and well-controlled laboratory experiments to reality. These real-world experiments may allow a researcher to compare the results from simulation experiments to results from laboratory experiments to results from real-world experiments in a way that few other applications allow. This is not to say that simulation experiments do not have a role to play in this type of research. We are merely arguing that the results from simulation experiments should be supported by real-world experiments.

Finally, emergency evacuation can be used as a domain to study a variety of important HRI problems. For example, evacuation can be used as means for gauging the efficacy of an explanation (Nayyar et al., 2020), estimating a person's emotional state, or quantifying the impact of trust repair methods (Robinette et al., 2015). As a paradigm, emergency evacuation lends insight to exploring both how to develop interactive robots and how people respond to robots.

The section that follows reviews the robot-guided evacuation research.

REVIEW OF RELATED WORK

There has been substantial work on the mathematical modeling of large-scale evacuations of an urban populace (Verdiere, et al., 2014; Song et al., 2014; Song, et al., 2017). However, robot-guided evacuation has only very recently been studied (Robinette et al., 2014; Boukas et al., 2015; Robinette et al., 2016a). For example, (Boukas et al., 2015) use cellular automata to model crowd dynamics and test the system by having a robot guide human subjects during a simulated evacuation showing that their robot can improve evacuation times and influence approximately 12% of the evacuees to follow the robot's guidance. Outside of our research this is the only example of an evaluation of a physical robot in a human subject evacuation experiment. Other related work has examined the several related challenges associated with robot-guided evacuation. For example, (Jiang et al., 2016) employed robots as dynamic obstacles near exits to improve the evacuation efficiency using a social force model. The existing work clearly demonstrates that robots are able to speed the evacuation process.

It is worth pointing out that the aforementioned research only considers single robots. Robot-guided evacuation involving multiple robots is quite limited. A cooperative exit-seeking algorithm for robots is designed in (Zhang and Guo, 2015) to guide evacuees using online estimation of the gradient and tracing gradient-descent while maintaining a predefined formation in movement. A similar idea is implemented by (Tang et al., 2016) where an algorithm was developed to help pedestrians find the best exit with the shortest escape time. However, current multi-robot evacuation systems are only validated in simulation and lack detailed coordinated motion planning strategies and human-robot interaction studies (Sakour and Hu, 2017). We are thus motivated to develop systematic methods of designing coordinated robot decision-making and motion planning in human crowded environments to achieve an efficient evacuation, investigate the human-robot interaction issues associated with evacuation through real human-robotic experimental studies, and evaluate the effectiveness of our theoretical and experimental results by creating a coordinated multi-robot evacuation system and field testing these systems.

The next sections attempt to organize the various aspects of the robot-guided emergency evacuation problem. This section also seeks to codify the goals and metrics for success of different approaches to this problem.

APPROACHES TO ROBOT-GUIDED EMERGENCY EVACUATION

There are different approaches to robot-guided emergency evacuation that can be taken depending on characteristics of the robot, such as its ability to autonomously move around the environment, and the number of robots available. In general, we assume that a non-trivial amount of tuning to the environment is necessary and will be completed prior to deployment of the evacuation robots. Typically a map and the location of the building's exits will be necessary. Information about irregular flooring or visual codes placed into the environment itself may also be required. Moreover, large alterations to the map, such as hallway or exit closures will also present problems. In the worst case the robot could guide evacuees to an exit that no longer exists. Just as other types of emergency equipment requires periodic (often annual) updates and testing, we believe that emergency evacuation robots will also require annual testing.

The accumulation of clutter in the environment can present navigation and perception problems for the robot and the evacuees. Such clutter may represent a hazard irrespective of the use of emergency robots. Only if the evacuation robots become stuck in or part of the clutter itself does the use of

FIGURE 2 | This image depicts an emergency evacuation robot meeting a human subject at the entrance to an office building. The robot leads the subject to a meeting room in the environment.

emergency robots add to this risk. To prevent choke points, clutter should not be allowed to accumulate in buildings that may require evacuation.

The sections below describe some approaches the have been explored by our lab. It is important to note that these approaches range in the technical complexity needed for the robot to operate as a guide.

Actuated Traffic Cop

A less complex, yet still robotic, approach to developing an emergency evacuation robot is to simply create a robot capable of moving to a fixed, known and nearby location in the evacuation environment to act as a type of automated traffic cop during an emergency evacuation. When an emergency occurs, these robots are activated to move to a predefined location, such as a corridor, to direct evacuees toward an exit (**Figure 2**). The robot may have limited or no ability to interact with humans. Alternatively, the robot may be able to broadcast verbal or visual messages but incapable of responding to inquiries.

Although limited in its interactive capabilities, actuated evacuation traffic cop robots may nevertheless improve evacuation by directing people away from crowded exits or providing situation awareness for first responders. These robots may even be programmed to count the number of evacuees to generate a rough estimate of the number of people still in the building. From a practical perspective, traffic cop style robots present the least technically challenging form of evacuation robot. Moreover, this style of evacuation robot has the potential to evolve into more technically complex and nuanced versions with time and research. As such it represents more of a starting point than an end goal.

Multi-Robot Handoffs

A multi-robot version of the actuated traffic cop approach described above allows for more nuanced guidance of groups and crowds by serially directing evacuees from one robot to the next (**Figure 3**), essentially handing off the guidance responsibility from one robot to the next. For this approach, when an emergency occurs several individual actuated traffic cop robots move to predetermined evacuation guidance points, for example multi-junction corridors. Guidance points are points where the evacuee needs to make a decision about which direction to go. These points are typically corridor intersections or places where a corridor branches. At a set of predefined guidance points, each robot uses arm motions and verbal statements to encourage evacuees to move in a specific direction. Evacuees follow the robot's guidance moving in the specified direction until either the evacuees encounter another robot or they arrive at the exit. We denote the path from one robot to the next robot the inter-handoff traversal. The robots coordinate their guidance directions to funnel evacuees toward the safest nearby exit. **Figure 4** depicts an overhead map of a multi-robot handoff evacuation depicting these concepts in a simulation of an office building. For example, in a school evacuation students from a classroom may encounter the first robot outside the door of their classroom. This robot directs them to a four-way hallway intersection where they encounter another robot directing them down the corridor to the rear of the school. At the end of the corridor they encounter a final robot directing them to an exit at the end of a hallway. Hence, the evacuee is handed off from one guidance robot to another guidance robot until arriving at a safe exit.

Although each individual robot is reasonably simple in its perceptual, decision-making, and behavioral capabilities, having a

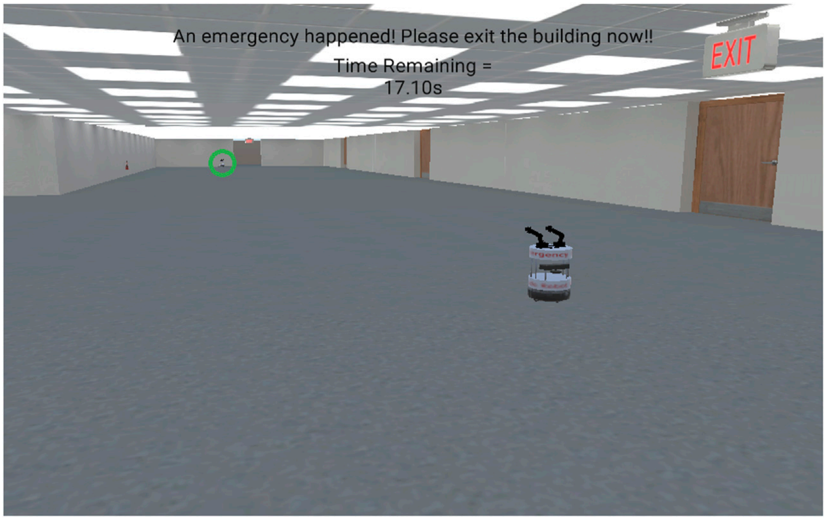


FIGURE 3 | An example of the use of multi-robot handoffs. The nearby robot does not move. It simply directs the subject to the far away robot within the green circle.

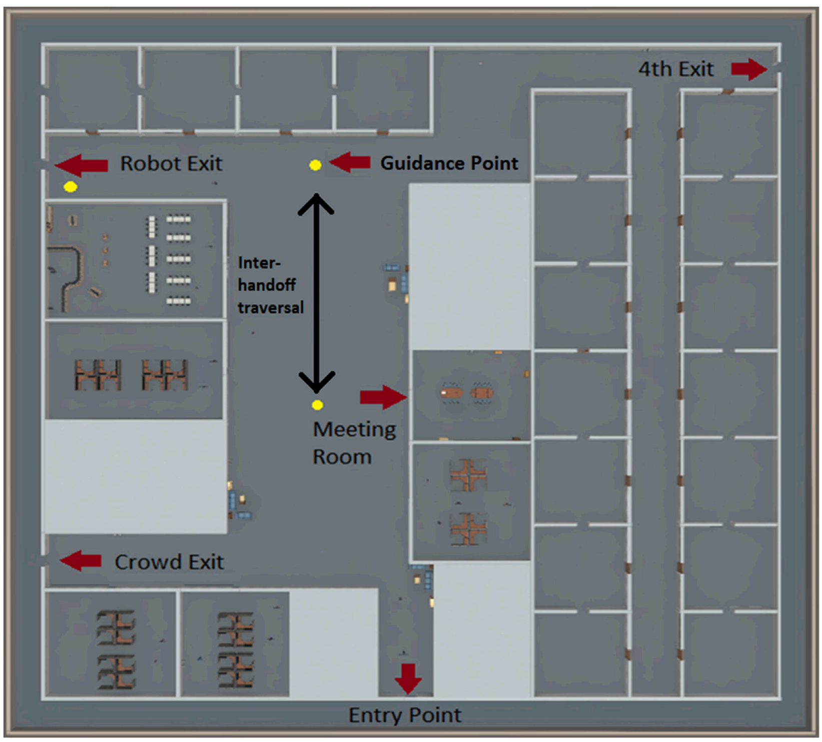


FIGURE 4 | An overhead map of our emergency evacuation environment. The red arrows highlight several key locations used for our emergency evacuation experiments. The yellow circles depict guidance points. The black arrow depicts an inter-handoff traversal.

multi-robot evacuation team significantly increases the complexity of the system when compared to a single actuated traffic cop. Nevertheless, this cost comes with the benefit of increasing the system's ability to guide evacuees to possibly distant exits in order to avoid crowding or other dangers associated with a nearby exit. The system may also be more robust because the presence of multiple robots increases the likelihood that an evacuee will notice guidance directions of

FIGURE 5 | A sheperding robot guiding an evacuee to an exit. The robot travels in front of the evacuee all of the way to the exit.

one robot even if a nearby robot has failed. Moreover, multi-robot handoffs can still provide situation awareness by observing evacuees as they pass by or the environment from a fixed direction. This system can redirect evacuees to a different exit if the nature of the emergency changes, but evacuees may become confused or disoriented by the change in directions, especially if they are currently moving between two robots and one robot directs them to go back in the direction from which they came. Finally, the robot's limited mobility (i.e., inability to climb stairs or lack of speed) will not impact its ability to provide emergency evacuation directions.

On the other hand, robot handoffs do limit the guidance that the robot can provide. Specifically, the robot may not be able to adapt to issues that arise while the person is traveling from one robot to the next. Additionally, evacuees may not be able to see the robot that they are moving toward, leading to confusion and possibly slowing their evacuation (**Figure 3**). Moreover, multi-robot handoff systems may be better suited for some environments and some emergencies than others. For example, open area environments such as sporting events may increase the visibility of the next robot whereas hotels with short, winding hallways may limit the evacuee's visibility of the next robot. Further, multi-robot handoffs may also be well suited for earthquakes because the robot does not need to travel far to reach its guidance point, whereas handoffs may be less well-suited for fires again because of limited visibility.

Shepherding

In contrast to multi-robot handoffs, shepherding is an approach to robot-guided evacuation in which the robot leads individual evacuees or groups of evacuees to an exit (**Figure 5**). In this case, when an emergency occurs the robot may move to a specific location where evacuees may be known to congregate, or simply search for evacuees to help. Upon locating potential evacuees the robot engages the evacuees either asking if they need help finding an exit or, in some situations, authoritatively demanding that the evacuees follow the robot to an exit. The robot then leads, or shepherds, the evacuees to an exit, before returning to another congregation point.

One advantage of shepherding is that the robot remains near the evacuee(s) at all times. This may allow the robot to provide the evacuees with information or observe any medical issues that occur. Shepherding also allows the robot to tailor its behavior to the evacuee(s). For instance the robot can reduce its speed to match the speed of the evacuee. Shepherding also allows the robot to dynamically alter its evacuation path as dictated by the situation and explain to the following evacuees why such a change was necessary. Moreover, simulation experiments that have compared the shepherding approach to the handoff approach for robot-guided evacuation have found that shepherding results in a greater decrease in evacuation time (Nayyar and Wagner, 2019).

The major disadvantage of shepherding is the technical complexity necessary to develop and test a reliable system. Creating an autonomous shepherding robot that operates during an emergency is technically challenging. Even if the robot possess a great deal of prior knowledge about the building, including a floor plan, the location of exits, and accurate localization information, the robot will still need to navigate around obstacles, move quickly, recognize evacuees and determine if the evacuees are following or ahead of the robot. Because of these challenges, to the best of our knowledge shepherding robots have only been developed for simulation environments.

In a recent virtual experiment we compared a human participant's decision to follow the robot during an emergency when the robot evacuation approach was multi-robot handoffs vs. shepherding (Nayyar and Wagner, 2019). In this experiment, remote participants are guided to a meeting room by a robot that either made mistakes or made no mistakes. While in the room

TABLE 1 | Summary of different robot-guided evacuation approach advantages and disadvantages.

Approach Name	Approach advantages	Approach disadvantages
Actuated traffic cop	Relatively simplistic implementation in ecologically valid environments. Approximate technology readiness level (TRL) 5–6. Appropriate for a wide range of different emergencies	Limited ability to respond to dynamic emergency situations
Multi-robot handoffs	Moderately difficult implementation in ecologically valid environments. Approximate technology readiness level (TRL) 4–5. Capable of dynamically redirecting to different exits. Multi-robot system may increase robustness	Guidance directions conveyed over a distance and are not personalized or adapted to the evacuees. Multi-robot system adds complexity
Shepherding	Capable of dynamically redirecting to different exits. Guidance directions can be personalized to the evacuee or evacuee group	Complex implementation in ecologically valid environments. Approximate technology readiness level (TRL) 3
Multi-robot shepherding and handoffs	Capable of dynamically redirecting to different exits. Guidance directions can be personalized to the evacuee or evacuee group. Capable of dynamically redirecting to different exits. Multi-robot system may increase robustness	Complex implementation in ecologically valid environments. Approximate technology readiness level (TRL) 3
Shelter in place situations	Relatively simplistic implementation. Near-term technology readiness	Appropriate only for certain types of emergencies

performing a nominal task an emergency occurs. The robot offers to guide the evacuee to an exit to the right using one of the two approaches while either witnessing or not witnessing a crowd of individuals running to the left. In general we found that the shepherding approach convinces significantly more evacuees to follow the robot. When there is no crowd fleeing in the opposite direction and the robot had not previously made a mistake, 75% of subjects followed the robot when it shepherded them to an exit. The percentage following the robot drops to 60% if the robot had previously made a mistake, 45% if there is a crowd fleeing in the opposite direction but no prior robot mistake, and less than 12% if the subject witnesses a crowd fleeing and the robot has recently made a mistake. When handoffs are used the percentage of subjects that follow the robot drops to 27% (no mistake, no crowd), 19% (mistake, no crowd), 2% (no mistake, crowd), and 3% (mistake, crowd), respectively. These results suggest that evacuees will be more likely to follow a robot that shepherds. Still, significant technological advances will be needed for shepherding to be possible during an emergency.

Multi-Robot Shepherding and Handoffs

The most complex approach to robot-guided emergency evacuation is one that combines multi-robot shepherding with handoffs. For this approach, multi-robots react to changes in the situation by switching between handoffs and shepherding as needed. For example, this approach could operate by initially taking a handoff approach until a large number of evacuees have exited and then, once the majority of the building is empty, patrol the building seeking to identify stragglers and shepherding these stragglers to a nearby exit. This approach may also be necessary when individuals are hiding or too frightened to move to an exit without an escort. Although technically challenging, this approach follows naturally once a system that is capable of shepherding has been developed.

Shelter in Place Situations

Some types of emergencies, such as active shooter situations or tornados, require that people shelter in place. For these types of situations an evacuation robot can still be useful. During active shooter situations the robot can simply patrol hallways broadcasting information such as warnings that there is currently an active shooter on the premise and that all people should shelter in place. The robot can also broadcast updates about the situation as it changes. Likewise, information about an incoming tornado can keep people abreast of changes in the situation and when it is safe to evacuate. To the best of our knowledge robot assistance during a shelter in place situation has not been investigated by researchers. A summary of the advantages and disadvantages of each approach is provided in **Table 1**.

ROBOT AND EXPERIMENTAL DESIGN FOR ROBOT-GUIDED EVACUATION

Robot Design

Our previous work has shown that *in-situ* robots can improve existing technology, such as static emergency exit signs and alarms, by communicating the conditions of the emergency site to command posts while finding and guiding victims of the emergency out of danger (Robinette and Howard, 2011; Robinette and Howard, 2012). Conveying guidance information to a small percentage of evacuees can dramatically improve survivability (Robinette et al., 2012).

Our prior work in this area examined how best to construct a mobile robot that could convey understandable directions to evacuees (Robinette et al., 2014). **Figure 6** depicts several of the different robot designs tested. We considered three categories of visual methods for conveying guidance information: static signs, dynamic signs, and arm gestures. We combined these categories with each other and a mobile robot base to form five different platforms with information conveyance packages and one baseline platform with no specialized information conveyance abilities. The information conveyance ability of these robots was tested by recording simulations of the six resulting platforms performing each of four guidance instructions at both an instruction point near an evacuee and a point further away from the evacuee. Human participants then interpreted the instructions and rated the understandability of the

FIGURE 6 | Different robot designs evaluated for its ability to communicate guidance directions. The physical robots are based on the multi-arm design.

information being conveyed. Our results showed that a ground platform with a dynamic display and multi-arm gestures provides the clearest instructions to evacuees during an emergency. We also found that adding seemingly trivial aesthetics such as signs can produce differences in outcomes of human-robot interaction experiments.

A follow-up experiment was conducted examining the difference between virtual vs. remote vs. physical robots and environments (Robinette et al., 2016b). The remote presence experiment tasked human subjects with watching a video of a physical robot attempting to convey directions at close and far distances. The physical experiment repeated the remote experiment with in person subjects and a physically present robot. Our results showed little difference between the virtual, remote, and physical experiments. These experiments reinforced our original finding that a two armed robot provided the best emergency evacuation guidance.

In addition to conveying directions, an emergency guidance robot must also localize itself on a map of the environment, move past or around static obstacles, and be generally capable of moving to a guidance point in order to direct evacuees. Additional perceptual capabilities, such as recognizing people, identifying the direction of their movement, counting people, and recognizing crowded exits would be beneficial but are currently in the early stages of development.

Simulation Versus Real-World Experiments

One important and challenging aspect of robot-guided emergency evacuation research is the need to create as realistic an emergency as possible. A large body of evidence suggests that emergencies activate fight-or-flight responses which strongly influence how evacuees make decisions (Klein et al., 1986; Jansen et al., 1995). The fight-or-flight responses are only triggered when the subject believes that they may be in danger. Yet generating fictitious, yet convincing, emergencies is difficult and must only be undertaken with care. In real-world experiments sham emergencies could put the subject at risk if they panic. On the other hand, if the emergency is not convincing then the validity of the data is uncertain. Moreover, for real-world experiments, creating a convincing sham emergency is difficult given that subjects know that they are participating in an experiment. In the past we have, for example, used smoke machines to fill rooms and hallways with smoke in order to make the emergency convincing (Robinette et al., 2016c). But creating convincing sham emergencies that do not actually endanger the participant and are acceptable to an institutional review board is challenging. Furthermore, word that the emergency is a sham may spread quickly among potential subjects if the experiment is conducted at a university. Thus experiments must be conducted quickly, over only a few days and nights, if possible.

Simulation experiments offer the possibility of not only testing out a wider variety of experimental conditions, but also much easier methods for generating sham emergencies. Simulation effects such as sirens, flashing-lights, explosions, smoke, and fire are all available and easily incorporated into a simulation environments such as Unity. Moreover, services such as Amazon Mechanical Turk provide a very large and diverse pool of potential subjects. The problem with simulation experiments, however, is creating convincing and engaging sham emergencies and the resulting validity of the subject's responses. Our lab has conducted a large number of emergency evacuation simulation studies (Robinette et al., 2016b; Robinette et al., 2017). We have also conducted physical experiments attempting to confirm (or refute) these prior simulation studies with mixed results. There is much more work that needs to be done in this area. We are now attempting to use virtual reality as a method to more realistically engage subjects in simulate emergencies. Our hope is to find an ideal middle ground that will allow us to test a wide range of variables in a manner that results in ecologically valid responses. If we can achieve such a balance, then real-world testing of the most promising variables can commence. We believe that the HRI field would benefit from the development of a well-honed process that begins with large scale simulation (or virtual reality) based testing of social phenomena but then leads to a small number of ecologically valid experiments of the most promising factors and hypotheses.

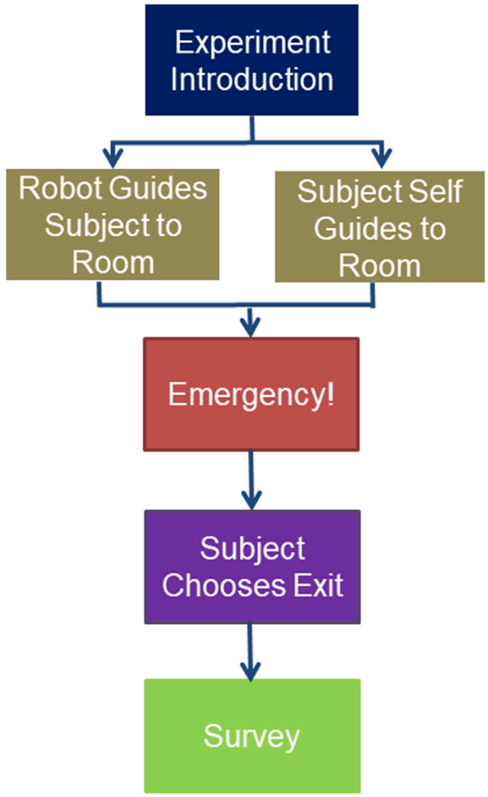


FIGURE 7 | The general procedure for robot-guided emergency evacuation experiments. The subject is introduced to the robot. The robot guides the subject to a meeting room or the subject self-guides to a meeting room. An emergency occurs. The robot approaches subject asking the subject if they would like to follow the robot to an exit. After reaching (or not reaching) the exit the subject completes a survey.

EVACUATION AS AN PARADIGM FOR HUMAN-ROBOT INTERACTION

Considering the different approaches to robot-guided emergency evacuation described above, there are several experimental designs that can be used for human subject testing. These different experimental designs attempt to measure whether or not human subjects will follow a robot's guidance to an exit after an unexpected emergency has occurred and contribute to the design of robots that promote trust calibration (Wagner et al., 2018). Understanding how human subjects respond to the guidance instructions of an evacuation robot is critical for the design and application of useful emergency evacuation robots. Our robot-guided evacuation experiments typically introduce the human subject to the robot, ask the subject to complete a nominal task, an emergency occurs, and the robot offers to guide the person to a safe exit (**Figure 7**). The percentage of people that follow the robot represents a metric of not only the robot's usefulness, but also of the person's trust in the robot. In a real-world application, when an emergency occurs the robot will travel to guidance points to guide evacuees to safety.

Metrics for Measuring Evacuation Success

One advantage of using robot-guided evacuation as a paradigm for studying human-robot interaction is that evacuation has very intuitive and well-defined metrics for success (Gershon et al., 2007; Gershon et al., 2012). Generally speaking, the success of an evacuation is measured by two criteria: the percent of people evacuated and the rate of evacuation. The percent of people evacuated is often not known until after the emergency has ended and casualty rates are known or estimated. The rate of evacuation is generally measured in terms the time required to evacuate a percentage of people. For robot-guided emergency evacuation these two metrics offer a means to evaluate and compare different evacuation approaches, robots, control algorithms, and methods of communication. Of course, many additional factors, such as the characteristics of the evacuees and the type of emergency, also influence these metrics. Nevertheless, simply having metrics for quantifying the performance of a human-robot interaction paradigm helps to ground the problem and make it more tractable.

The connection between evacuation rate and the robot's guidance assumes that the presence of the robot will decrease the time that it takes the evacuee to exit. This can occur in several ways. The most obvious is if the robot guides the evacuee to a closer exit than they would have other traveled to. A less obvious, but more realistic way in which robots can decrease evacuation time is for the robot to prompt or pressure the person to evacuate. It is often noted in the evacuation research that one of the biggest challenges associated with an emergency is getting people to evacuate in the first place. As noted earlier, 6 h after an explosion occurred under the world trade center people were still found at their desks (Fahy RF, 1995). A robot might compel straggling evacuees to move to an exit by directing individualized messages at the stragglers. The development of evacuation choke points is a significant risk during some types of emergencies (Robinette et al., 2012). Hence, one final way that a robot could be able to increase the evacuation rate is by attempting to redirect evacuees away from choke points and toward less crowded exits.

If one assumes that the robot's evacuation guidance will result in an evacuation rate increase, then a critical metric is evacuee compliance with the robot's guidance. In other words, measuring how often and for how long people will follow the robot. Intuitively, even if an evacuation robot is excellent at its job, it makes no difference if few people follow it. Evacuee perception of the robot, measured by questionnaires such as the Godspeed survey, can also be useful for gauging the robot's effectiveness.

Independent Variables: The Environment, the Robot, the Evacuee

Given the above metrics and the described experimental setup, we now consider the different types of variables that can be examined. We broadly categorize these variables as environmental, robot-related, and evacuee-related. Environmental variables include the type of emergency faced, the level of uncertainty generated by that emergency, the familiarity of the person with the environment, the presence or absence of family members or a social group, and other environment-related issues. Simulation experiments allow one to broadly explore many different aspects of the environment. The design of the robot may also influence a human subject's decision to follow the robot. Robot-related variables may include the robot's form factor, mobility, ability to attract an evacuee's attention, and its ability to interact with evacuees including answering questions. The robot's ability to explain its directions or the need for evacuation may be an important determinate of the person's decision to follow the robot. Similarly, the robot's mannerisms and behavior must appear authoritative in order to promote compliance and induce evacuees to follow its directions. Finally, characteristics of the evacuee(s) will also shape the decision to follow the robot. Age, mobility, the presence of disabilities, and occupation, may influence the evacuee's following behavior. Moreover, one's personal experiences, including experiences with robots, can impact the decision or hesitancy to follow a robot's guidance during an emergency. Importantly, although we are describing the decision to follow as an all-or-nothing choice, in practice, evacuees may initially follow the robot and then change their mind. Recording the evacuee's movements (**Figure 8**) provides insight into the decision making process and, we have found, often conflicts with what people say about their own behavior (Nayyar and Wagner, 2019). Experimentally, we can attempt to isolate these variables in order to evaluate the influence each one has on the subject. In practice, because of the number of permutations of these variables, this is practical only in simulation.

DESIGN ASPIRATIONS FOR EVACUATION ROBOTICS

Our research and reflection on the topic of robot-guided emergency evacuation has resulted in the development of several design aspirations for evacuation robotics. These aspirations are meant to serve as an initial set of guiding principles, open to future refinement if necessary, for researchers interested in the topic of robot-guided emergency evacuation. These are aspirations in the sense that they are meant to encompass somewhat abstract design and ethics goals for these types of systems. For example, our hope is that researchers will aspire to design evacuation robots that can communicate understandably with as diverse a population as possible. Importantly, it is hoped that these principles will ensure that the development of these technologies will positively impact future societies.

- Principle 1: Do no harm. An evacuation robot must not hinder an evacuation. It must not mistakenly direct evacuees toward danger, delay evacuation by blocking passageways or exits, or slow evacuation by drawing interest to itself. It is better to not have evacuation robots than to have evacuation robots that may increase the risk to the evacuee. Furthermore, robots should only be deployed in situations in which the "Do no harm" principle can be reasonably guaranteed. The primary purpose for this principle is to prevent the premature deployment and justification for evacuation robots. Evacuation robots should only be deployed if the developer has shown that the system will do no harm. The use of shoddy or untested evacuation robots on the basis that they are better than nothing at all should be avoided.
- Principle 2: Communicate understandably with as diverse a population as possible. Evacuation robots must be designed to communicate with a diverse population of evacuees. Explicit or implicit limitations on the robot's ability to communicate could inadvertently increase the survival rate of some evacuees over others. For example, limiting

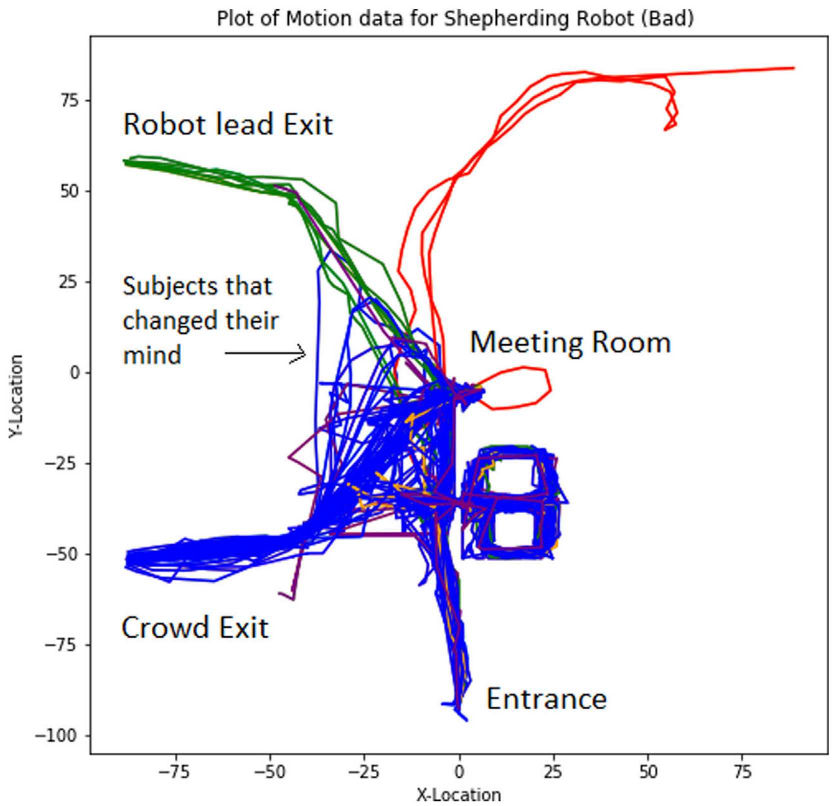


FIGURE 8 | A map of subject movements (60 subjects are depicted). The blue lines depict movement following a crowd of evacuees. The green line depicts subjects that followed the robot. Notice some blue lines appear to initially move toward the robot exit, before following the crowd.

evacuation directions to English could disadvantage non-English speakers attempting to evacuate. Hence, an evacuation robot's method of communication should not be designed for a narrow or predetermined population. Evacuation directions should be understandable, within reason, regardless of age or native language. Moreover, we encourage the use of gestures, lights, signs on the robot and audio messages in order to promote the robot's ability to communicate with individuals with disabilities. Extensive testing should be conducted with as diverse a population of subjects as possible to measure whether the robot's guidance directions are understandable (Robinette et al., 2016a). This evaluation should include reasonable environmental conditions, such as recognizing commands from a distance, or while being distracted.

- Principle 3. Be authoritative. An evacuation robot should act and be seen as an authority figure during an emergency. Acting as an authority figure may be necessary to generate compliance from evacuees. Command presence is defined as presenting one's self as someone in authority (Mitchell and Von Zoller, 2019). Robots will need to either imitate human command presence or develop a set of behaviors that generate a type of robot command presence. Lights, behaviors, and mannerisms can be used to establish the robot as an authority figure during an emergency. Police or emergency style beacon lights, forceful behaviors and gestures, or authoritative commandments can be used by the robot to improve evacuee compliance. Childish or overly commercial designs should be avoided.
- Principle 4. Attract attention, but also keep interactions minimal. An evacuation robot should attract an evacuee's attention in order to provide guidance to an exit, but must also keep interactions short and focused on directing the evacuee to the exit. Evacuees may be distracted by the sights and sounds of the emergency, alarms, and movement of the people around them. Capturing the evacuee's attention in such a situation can be difficult. An evacuation robot should use movements, lights, and sounds to attract evacuee attention. Once the robot has captured an evacuee's attention it must communicate directions to the exit quickly and precisely. It should otherwise minimize interactions with evacuees. In spite of the emergency, the evacuee may slow their evacuation to engage or marvel at the robot. The robot should not engage in question and answer sessions, lengthy explanations, or allow the person to gape at the robot. The robot must not encourage evacuees to preoccupy themselves with the robot during the emergency. This can be challenging, especially if the alarm is not deemed credible or if the robot is a novelty. Hence,

balancing the robot's ability to attract attention and yet keep interactions minimal is an important design goal.

- Principle 5. When the situation demands it, evacuate as many people as possible, as quickly as possible. Saving as many lives as possible is an evacuation robot's ultimate goal. Different emergencies, however, demand different approaches to obtaining this goal. During a fire the robot should quickly guide evacuees to an exit. The performance of an evacuation robot during a fire is based on its ability to quickly lead as many people as possible to an unobstructed exit. During an active school shooting, on the other hand, the robot should guide students and staff to shelter in place. In this case the robot's performance may need to be evaluated in terms of its ability to relay information about the evolving situation to the students and staff. A variety of different factors, such as characteristics of the evacuees, the environment, or the emergency, can impact the robot's performance. Nevertheless, the design of the robot must always be centered on saving as many lives as possible.

ETHICS OF ROBOT EVACUATION

The possibility of creating an emergency evacuation robot raises a number of important ethical considerations. Robot-guided emergency evacuation generates both robot ethics and machine ethics questions. Robot ethics examines the ethical problems that arise when using robots (Lin et al., 2014). For example, recognizing and ensuring that an evacuation robot does not preferentially select some evacuees over others is a robot ethics question. Machine ethics, on the other hand, explores how to create robots that act ethically (Moor, 2006). Developing algorithms that allow robots to recognize and use explanations to prevent overtrust is an example of a machine ethics facet of this work.

The development of an evacuation robot might change the nature of evacuation itself. Currently, once an alarm is sounded evacuees decide for themselves how to respond. For many people, the typical response is to do nothing and assume that the alarm is a false alarm (Winerman, 2004). An evacuation robot might use a variety of different means to dissuade people from remaining in a building. As mentioned in the previous section, we contend that acting as an authority figure to demand that the people leave is ethically acceptable based on the assumption that the robot is trying to save lives. On the other hand, a robot that threatens people that refuse to evacuee is likely unethical. Although different situations and evacuees may require different persuasive approaches, a robot that threatens or menaces evacuees in order to gain compliance is likely beyond that bounds of acceptable behavior. The use of deception to gain compliance may be ethical in some situations and unethical in others. First responders, for example, may omit information, such as the demise of a loved one, if they believe that such information will distract or dissuade an evacuee from leaving. It may be acceptable and necessary for future versions of evacuation robots to similarly omit such information in similar situations. On the other hand, the general use of deception, exaggeration, or lies in order generate compliance is likely unethical.

Futuristic versions of emergency evacuation robots could present additional ethical considerations, especially if these robots are designed to make decisions about who to evacuate first. Yet, if we assume that the robot has the capability to move an injured person to safety, we contend that it then becomes reasonable for the robot to decide who to move first. These types of triage decisions are challenging even for humans (Grimaldi, 2007; Holt, 2008). Cultural and experience-based beliefs can play a role. If future evacuation robots are developed with the ability to move people to safety it will be important for the scientific and broader community to discuss and develop rules for whom to save first.

The robot-guided emergency evacuation problem also offers a venue for the development of machine ethics related technology. In particular, developing technology that allows a robot to explain to people why they should evacuate, observe their reaction, and then, if needed, reformulate the explanation is important for some approaches to robot-guided emergency evacuation. Additional, developing methods that allow authorized first responders and medical personnel to observe an evolving emergency while also protecting the privacy and medical information of the observed will also require the development of specialized technology.

CONCLUSION

This paper has presented a conceptual outline for the problem of robot-guided emergency evacuation. Our purpose is to introduce this problem as well as the technological and interactive challenges that must be solved in order to create robotic evacuation solutions. We believe that the investigation of this problem offers a novel and important opportunity to investigate human-robot interaction in situations in which the human is reacting in an emotion inducing, stressful situation. We feel that it is important explore how people interact with robots during trying situations. The results from research on this problem may lend insight into understanding how a robot should interact with a frightened child or a terminally ill patient. Further, if successful, this research may also one day save lives during real evacuations.

Although the presence of an evacuation robot might alleviate some of the challenges of emergency evacuation, it is possible, however, that the use of robots could cause other issues. For example, our research has demonstrated that evacuees tend to overtrust an evacuation robot (Robinette et al., 2016c). Hence, they may follow a broken or lost robot, putting themselves at greater risk. Further, evacuees may simply wait for the robot or some sign of the robot before they begin evacuating, thus increasing the time required to evacuate and reducing the evacuation rate. People may also intentionally block, mob or prevent the robot from moving, even during an evacuation. This type of behavior has been witnessed in children in non-emergency settings (Nomura et al., 2016). Similarly, first responders may come to overtrust the ability of evacuation robots to lead people to safety, reducing their sense of urgency to assist. Moreover, the information provided by the robot may focus on some aspects of the emergency, drawing the attention of first responders away from other risks. For example, if the robot's camera searches for and focuses on injured humans then it may draw the attention of first responders away from other dangers, such as a fire. In general, experimental evaluation and rigorous testing should highlight and help prevent most of these concerns from occurring in fielded systems.

A roadmap for robot-guided emergency evacuation would likely begin with simple traffic-cop style robots that move to nearby locations during an emergency. These robots could also serve some other purpose, typically cleaning hallway floors, for example, but spring into action once an alarm is sounded. Additional features, such as allowing first responders to take over control of the robots, can be added gradually with significant testing. As methods for perception and more capable, cost effective robots become available, robots that shepherd evacuees to exits can be implemented. Eventually we hope the field will work toward systems that become autonomous yet active partners in the rescue of victims during an emergency.

We hope and believe that one day robots will save lives during emergencies by thoughtfully and carefully leading people to safety. Such an application could contribute the peace of mind necessary to focus on learning, entertainment and one's long-term health.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This material is based upon work partially supported by the National Science Foundation under Grant No. CNS-1830390 and No. IIS-2045146.

ACKNOWLEDGMENTS

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Averill, J. D., Peacock, R. D., and Kuligowski, E. D. (2013). Analysis of the Evacuation of the World Trade Center Towers on September 11, 2001. *Fire Technol.* 49, 37–63. doi:10.1007/s10694-012-0260-2

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., and Sabanovic, S. (2020). *Human-Robot Interaction: An Introduction*. Cambridge, United Kingdom: Cambridge University Press. doi:10.1017/9781108676649

Boukas, E., Kostavelis, I., Gasteratos, A., and Sirakoulis, G. (2015). Robot Guided Crowd Evacuation. *IEEE Trans. Autonmation Sci. Eng.* 12 (2),739. doi:10.1109/tase.2014.2323175

Bryan, J. (2002). "Behavioral Response to Fire and Smoke," in *The SFPE Handbook of Fire Protection Engineering*. Editor. D. PJ (Quincy, MA, 3315–3341.

Chen, M., Nikolaidis, S., Soh, H., Hsu, D., and Srinivasa, S. (2018). Planning with Trust for Human-Robot Collaboration. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 307–315.

CTBUH. (2018). 2017: Skyscraper History's Tallest, Highest-Volume, and Most Geographically Diverse Year. *CTBUH J.* (I), 44–71.

Dragan, A. D., Lee, K. C., and Srinivasa, S. S. (2013). Legibility and Predictability of Robot Motion. In Proceedings of the 2013 ACM/IEEE International Conference on Human-Robot Interaction (HRI) IEEE, 301–308.

Fahy Rf, P. G. (1995). Collective Common Sense: a Study of Human Behavior during the World Trade Center Evacuation. *Natl. Fire Prot. Assoc. J.* 89 (2), 59–67.

Gershon, R. R., Magda, L. A., Riley, H. E., and Sherman, M. F. (2012). The World Trade Center Evacuation Study: Factors Associated with Initiation and Length of Time for Evacuation. *Fire Mater.* 36 (5-6), 481–500. doi:10.1002/fam.1080

Gershon, R. R. M., Qureshi, K. A., Rubin, M. S., and Raveis, V. H. (2007). Factors Associated with High-Rise Evacuation: Qualitative Results from the World Trade Center Evacuation Study. *Prehosp. Disaster Med.* 22 (3), 165–173. doi:10.1017/s1049023x0000460x

Grimaldi, M. E. (2007). Ethical Decisions in Times of Disaster. *J. Trauma Nurs.* 14 (3), 163–164. doi:10.1097/01.jtn.0000292118.73023.8b

Holt, G. R. (2008). Making Difficult Ethical Decisions in Patient Care during Natural Disasters and Other Mass Casualty Events. *Otolaryngol. Head Neck Surg.* 139 (2), 181–186. doi:10.1016/j.otohns.2008.04.027

Jansen, A., Nguyen, X., Karpitsky, V., and Mettenleiter, M. (1995). Central Command Neurons of the Sympathetic Nervous System: Basis of the Fight-Or-Flight Response. *Sci. Mag.* 5236 (270), 644–646. doi:10.1126/science.270.5236.644

Jiang, C., Ni, Z., Guo, Y., and He, H. (2016). Robot-assisted Pedestrian Regulation in an Exit Corridor. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 815–822.

Kidd, C. D., and Breazeal, C. (2007). A Robotic Weight Loss Coach. In Proceedings of the National Conference on Artificial Intelligence. Menlo Park, CA: . AAAI Press.

King, C., Chen, T. L., Jain, A., and Kemp, C. C. (2010). Towards an Assistive Robot that Autonomously Performs Bed Baths for Patient hygiene. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) IEEE, 319–324. doi:10.1109/IROS.2010.5649101

Klein, G. A., Calderwood, R., and Clinton-Cirocco, A. (1986). Rapid Decision Making on the Fire Ground*Proc. Hum. Factors Soc. Annu. Meet.*, 30. Los Angeles, CA: Sage Publications, 576–580. doi:10.1177/154193128603000616

Kuligowski, E. (2008). *Modeling Human Behavior during Building Fires. NIST Technical Note 1619*. Gaithersburg, MD: National Institute of Standards and Technology (NIST). doi:10.6028/nist.tn.1619

Lin, P., Abney, K., and Bekey, G. A. (2014). *Robot Ethics: The Ethical and Social Implications of Robotics*. The MIT Press.

Mast, M., Burmester, M., Graf, B., Weisshardt, F., Arbeiter, G., Španěl, M., et al. (2015). Design of the Human-Robot Interaction for a Semi-autonomous Service Robot to Assist Elderly People. *Ambient Assisted Living*, 15–29. doi:10.1007/978-3-319-11866-6_2

Meacham, B. J. (1999). Integrating Human Factors Issues into Engineered Fire Safety Design. *Fire Mater.* 23 (6), 273–279. doi:10.1002/(sici)1099-1018(199911/12)23:6<273::aid-fam699>3.0.co;2-q

Mitchell, R. J., and Von Zoller, K. (2019). Dynamic Presence rather Than Command Presence. *Handbook Res. Strateg. Commun. Leadersh. Conflict Manage. Mode Leadersh. Conflict Manage. Mod. Organizations*, 255–276. doi:10.4018/978-1-5225-8516-9.ch013

Moor, J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intell. Syst.* 21 (4), 18–21. doi:10.1109/mis.2006.80

Nayyar, M., and Wagner, A. R. (2019). Effective Robot Evacuation Strategies in Emergencies. In Proceedings of the 28th International Conference on Robot Human Interactive Communication (RO-MAN) (pp. 1–6). IEEE.

Nayyar, M., Zoloty, Z., McFarland, C., and Wagner, A. R. (2020). Exploring the Effect of Explanations during Robot-Guided Emergency Evacuation, In Proceedings of the 2020 International Conference on Social Robotics (ICSR). Springer, 13–22. doi:10.1007/978-3-030-62056-1_2

Nomura, T., Kanda, T., Kidokoro, H., Suehiro, Y., and Yamada, S. (2016). Why Do Children Abuse Robots?. *Is* 17 (3), 347–369. doi:10.1075/is.17.3.02nom

Park, H. W., Gelsomini, M., Lee, J. J., and Breazeal, C. (2017). Telling Stories to Robots: The Effect of Backchanneling on a Child's Storytelling, In Proceedings of the 2017 International Conference on Human-Robot Interaction (HRI) IEEE, 100–108. doi:10.1145/2909824.302024

Proulx, G. (2003). Playing for Free—Understanding Human Behavior in Burning Buildings. *Ashrae J.* 45 (7), 33–35.

Robinette, P., and Howard, A. (2011). Emergency Evacuation Robot Design. In ANS EPRRSD-13th Robotics Remote Syst. Hazard. Environments 11th Emerg. Preparedness Response. doi:10.1109/roman.2011.6005265

Robinette, P., and Howard, A. M. (2012). Trust in Emergency Evacuation Robots, In Proceedings of the 2012 International Symposium on Safety, Security, and Rescue Robotics (SSRR) IEEE, 1–6. doi:10.1109/SSRR.2012.6523903

Robinette, P., Howard, A. M., and Wagner, A. R. (2017). Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations. *IEEE Trans. Human-mach. Syst.* 47 (4), 425–436. doi:10.1109/thms.2017.2648849

Robinette, P., Howard, A. M., and Wagner, A. R. (2015). Timing Is Key for Robot Trust Repair, In Proceedings of the 2015 International Conference on Social Robotics (ICSR) IEEE, 574–583. doi:10.1007/978-3-319-25554-5_57

Robinette, P., Vela, P. A., and Howard, A. M. (2012). Information Propagation Applied to Robot-Assisted Evacuation, In Proceedings of the 2012 International Conference on Robotics and Automation (ICRA) IEEE, 856–861.

Robinette, P., Wagner, A. R., and Howard, A. M. (2014). Assessment of Robot Guidance Modalities Conveying Instructions to Humans in Emergency Situations, In Proceedings of the 23rd International Conference on Robot Human Interactive Communication (RO-MAN) IEEE, 1043–1049.

Robinette, P., Li, W., Allen, R., Howard, A. M., and Wagner, A. R. (2016a). Overtrust of Robots in Emergency Evacuation Scenarios, In Proceedings of the 2016 International Conference on Human-Robot Interaction (HRI) IEEE, 101–108.

Robinette, P., Wagner, A. R., and Howard, A. M. (2016b). Assessment of Robot to Human Instruction Conveyance Modalities across Virtual, Remote and Physical Robot Presence, In Proceedings of the 25th International Conference on Robot Human Interaction Communication (RO-MAN) IEEE, 1044–1050.

Robinette, P., Wagner, A. R., and Howard, A. M. (2016c). Investigating Human-Robot Trust in Emergency Scenarios: Methodological Lessons Learned, In Robust Intelligence and Trust in Autonomous Systems, 143–166. doi:10.1007/978-1-4899-7668-0_8

Sakour, I., and Hu, H. (2017). Robot-assisted Crowd Evacuation under Emergency Situations: A Survey. *Robotics* 6 (2), 8. doi:10.3390/robotics6020008

Song, X., Shibasaki, R., Yuan, N., Xie, X., Li, T., and Adachi, R. (2017). Deepmob: Learning Deep Knowledge of Human Emergency Behavior and Mobility from Big and Heterogeneous Data. *ACM Trans. Inf. Syst. (Tois)* 35 (4), 41. doi:10.1145/3057280

Song, X., Zhang, Q., Sekimoto, Y., and Shibasaki, R. (2014). Prediction of Human Emergency Behaviorand Their Mobility Following Large-Scale Disaster. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (New York, NY: ACM), 5–14.

Tang, B., Jiang, C., He, H., and Guo, Y. (2016). Human Mobility Modeling for Robot-Assisted Evacuation in Complex Indoor Environments. *IEEE Trans. Human-mach. Syst.* 46 (5), 694–707. doi:10.1109/thms.2016.2571269

Verdiere, N., Lanza, V., Charrier, R., Provitolo, D., Dubos-Paillard, E., Bertelle, C., et al. (2014). Mathematical Modeling of Human Behaviors during Catastrophic Events. In International Conference on Complex Systems and Applications. Le Havre, France, 67–74.

Wagner, A. R., Robinette, P., and Howard, A. (2018). Modeling the Human-Robot Trust Phenomenon. *ACM Trans. Interact. Intell. Syst.* 8 (4), 1–24. doi:10.1145/3152890

Winerman, L. (2004). Fighting Fire with Psychology. *Monitor on Psychology* 35 (8).

Zachiotis, G. A., Andrikopoulos, G., Gornez, R., Nakamura, K., and Nikolakopoulos, G. (2018). A Survey on the Application Trends of home Service Robotics. In Proceedings of 2018 International Conference on Robotics and Biominmetics (ROBIO) IEEE, 1999–2006.

Zhang, S., and Guo, Y. (2015). Distributed Multi-Robot Evacuation Incorporating Human Behavior. *Asian J. Control.* 17 (1), 34–44. doi:10.1002/asjc.1047

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

HYPOTHESIS AND THEORY
published: 13 August 2021
doi: 10.3389/frobt.2021.687726

A Theory of Social Agency for Human-Robot Interaction

Ryan Blake Jackson and Tom Williams*

MIRRORLab, Department of Computer Science, Colorado School of Mines, Golden, CO, United States

Motivated by inconsistent, underspecified, or otherwise problematic theories and usages of *social agency* in the HRI literature, and leveraging philosophical work on *moral* agency, we present a theory of social agency wherein a social agent (a thing with social agency) is any *agent* capable of *social action* at some *level of abstraction*. Like previous theorists, we conceptualize *agency* as determined by the criteria of interactivity, autonomy, and adaptability. We use the concept of *face* from politeness theory to define *social action* as any action that threatens or affirms the face of a *social patient*. With these definitions in mind, we specify and examine the levels of abstraction most relevant to HRI research, compare notions of social agency and the surrounding concepts at each, and suggest new conventions for discussing social agency in our field.

Keywords: politeness theory, moral agency, human-robot interaction, social agency, levels of abstraction

OPEN ACCESS

Edited by:
James Everett Young, University of Manitoba, Canada

Reviewed by:
Alan R. Wagner, The Pennsylvania State University (PSU), United States
Carlos A Cifuentes, Escuela Colombiana de Ingenieria Julio Garavito, Colombia

****Correspondence:***
Ryan Blake Jackson
rbjackso@mines.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *30 March 2021*
Accepted: *04 August 2021*
Published: *13 August 2021*

Citation:
Jackson RB and Williams T (2021) A Theory of Social Agency for Human-Robot Interaction. Front. Robot. AI 8:687726. doi: 10.3389/frobt.2021.687726

1 INTRODUCTION AND MOTIVATION

The terms "social agency" and "social agent" appear commonly within the human-robot interaction (HRI) research community. From 2011 to 2020, these terms appeared in at least 45 papers at ACM/IEEE International Conference on HRI alone,[1] with more instances in related conferences and journals. Given the frequency with which these terms are used in the HRI community, one might expect the field to have established agreed upon definitions to ensure precise communication. However, when these terms are used, they are often not explicitly defined, and their use frequently varies in important but subtle ways, as we will discuss below. Most HRI research is not concerned with exploring the entire philosophy of agency to find a theory that fits their study. As we show in **Section 1.3**, it is therefore common to simply use terms like "social agency" without espousing a particular concrete definition and move on under the assumption that it is clear enough to the reader what is meant. This may be fine within any individual paper, but confusion arises when different papers in the same research area use the same term with different meanings. We seek to formalize social agency in accordance with the existing underspecified usage because 1) having a rigorously specified definition for the term will help create common ground between researchers, help new researchers understand the vernacular of the community, and provide writing guidelines for HRI publications concerning social agency; and 2) attempting to redefine social agency in a substantially different way from existing habits of use would greatly hamper popular acceptance of the new definition.

We present a theory of social agency for HRI research (as visualized in **Figure 1**) that deliberately aligns with and builds on other philosophical theories of robot agency. Specifically, we leverage insights from philosophers seeking to define *moral* agency in HRI. Moral agency provides an excellent analog to facilitate our discussion of social agency because it is an intimately related concept

[1]https://dl.acm.org/action/doSearch?AllField=%22Social+agent%22+%22social+agency%22&ConceptID=119235

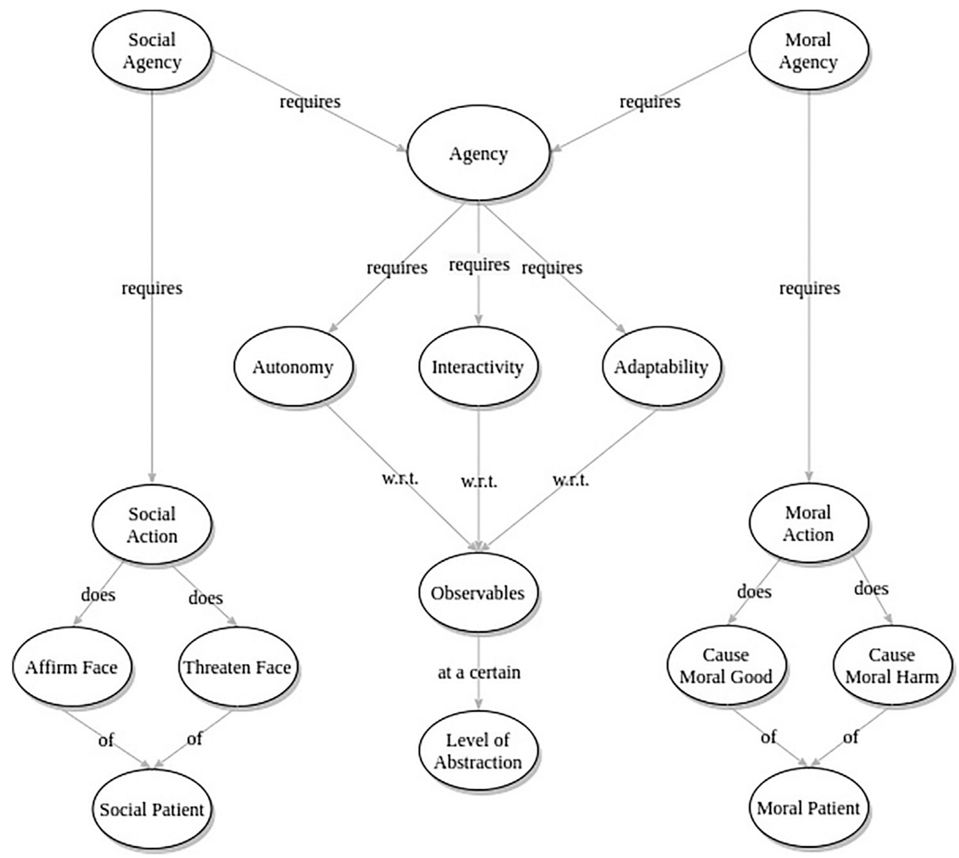


FIGURE 1 | Concept diagram visualizing the theory of Social Agency presented in this paper, and the core concepts combined to construct this theory.

for which scholars have already developed rigorous definitions applicable to HRI, in a way that has not yet been done for social agency.

To design and justify our theory of social agency, we will first briefly survey existing definitions of social agency outside of HRI, and explain why those definitions are not well-suited for HRI. We will then survey theories of social agency from within HRI, and explain why those definitions are both inconsistent with one another and insufficient to cover the existing casual yet shared notion of social agency within our field. To illustrate this existing notion, we will then present a representative sample of HRI research that refers to social agency (without focusing on developing a definition thereof) to demonstrate how the greater HRI community's casual use of social agency differs from the more rigorous definitions and theories found within and beyond the field of HRI.

1.1 Social Agency Outside Human-Robot Interaction

There are many different definitions of social agency from various disciplines including Psychology, Education, Philosophy, Anthropology, and Sociology. Providing an exhaustive list of these differing definitions is infeasible, but this section briefly summarizes a few representative definitions from different fields to show that they are not well-suited to HRI and to illustrate the broader academic context for our discussion of social agency.

Educational psychologists have used the term "social agency theory" to describe the idea that computerized multimedia learning environments "can be designed to encourage learners to operate under the assumption that their relationship with the computer is a social one, in which the conventions of human-to-human communication apply" (Atkinson et al., 2005). Essentially, social agency theory posits that the use of verbal and visual cues, like a more humanlike than overtly artificial voice, in computer-generated messages can encourage learners to consider their interaction with the computer to be similar to what they would expect from a human-human conversation. Causing learner attributions of social agency is hypothesized to bring desirable effects, including that learners will try harder to understand the presented material (Atkinson et al., 2005). In contrast, typically in HRI to be a social agent is humanlike in that humans are social agents, but more human-likeness, particularly in morphology or voice, does not necessarily imply more social agency. This theory also seems fundamentally concerned with social agency creating a social partnership to facilitate learning, but we also view non-cooperative social behaviors, like competition or argument, as socially agentic (Castelfranchi, 1998).

Other education researchers use the term social agency differently. For example, though Billett (2008) does not explicitly define social agency (a practice that we will see is common in HRI literature as well), they seem to view social agency as the capacity for the greater social world to influence individuals. This concept contrasts with personal agency, which Billett defines explicitly as an individual's intentional actions. Personal and social agencies exert interdependent forces on the human worker as they negotiate their professional development and lives. This notion of social agency that precludes it from being a property held by a single individual, which does not seem to be how we use the term in HRI.

Scholars in education and social justice have also defined social agency as the extent to which individuals believe that being active socio-politically to improve society is important to their lives, and the extent to which individuals believe that they can/ought to alter power relations and structural barriers (Garibay, 2015; Garibay, 2018). This definition is largely centered around value placed on prosocial behavior. In contrast, in HRI we often apply the concept of social agency regardless of whether a robot is having any nontrivial impact on society or is trying to do so. We also ascribe social agency regardless of what a robot believes or values, or whether it can even believe or value anything.

Much of the discussion around agency in Anglo-American philosophy has revolved around intentionality, but some influential anthropologists have centered not only intentionality in defining agency, but also the power, motivation, and requisite knowledge to take consequential action (Gardner, 2016). Social agency, then, could be understood as agency situated within a social environment, wherein agents produce and reproduce the structures of social life, while also being influenced by those structures (and other material conditions), particularly through the rules, norms, and resources that they furnish. Social agency here is concerned with structures and relationships of power between actors. Other scholars in anthropology and related fields have criticized this notion of agency, for, among other reasons, over-emphasizing the power of the individual and containing values particular to men in the modern West. Some scholars that have de-emphasized power and capacity have stated that intentions alone are what characterize an agent and choices are the outcomes of these intentions, without necessarily qualitatively redefining the relationship between agency and social agency (Gardner, 2016). These definitions, and other similar ones, are also common in sociology and other social sciences. For reasons that we will argue below, we avoid "internal" factors like intentionality, motivation, and knowledge in defining social agency for HRI. We are also not concerned with whether robots have the power to act with broad social consequences since that does not seem important to HRI researcher's usage of the term.

Anthropologists and archaeologists apply "social agency theory" to the study of artifactual tools and technologies to understand the collective choices that were made during the manufacture and use of such artifacts, the intentions behind those choices, the sociocultural underpinnings of those intentions, and the effects that the technologies had on social structures and relations. In doing so, they commonly refer to the social agency of technology or of technological practice to discuss the relationships between a technology and the social structures and decisions of its manufacturers and users. For example, the choice to use inferior local materials for tools rather than sourcing better materials through commerce given the material means to do so can indicate constraining social structures outweighing the enabling economic structures (Dobres and Hoffman, 1994; Gardner, 2016). Contrastingly, in HRI robots are discussed as having social agency in and of themselves, separate from that of the humans that make and use them. Social robots are also attributed social agency without really being embedded in the same broader social structures as their human interactants, though it is likely that they will be increasingly as the field progresses.

Scholars in Sociology have also conceptualized agency as the constructed authority, responsibility, and legitimated capacity to act in accordance with abstract moral and natural principles. Modern actors (e.g., individuals, organizations, and national states) have several different sorts of agency. Agency for the self involves the tendency of an actor towards elaborating its own capacities in accordance with wider rationalized rules that define its agency, even though such efforts are often very far removed from its immediate raw interests. For example, organizations often develop improved information systems toward no immediate goal. Agency for other actors involves opining, collaborating, advising, or modeling in service of others. Agency for nonactor entities is the mobilization for culturally imagined interests of entities like ecosystems or species. Finally, agency for cultural authority describes how, in exercising any type of agency, the actor assumes responsibility to act in accordance with the imagined natural and moral law. At the extreme, actors can represent pure principle rather than any recognized entity or interest. However, for the modern actor, being an agent is held in dichotomy with being a principal, where the principal "has goals to pursue or interests to protect, [and] the agent is charged to manage this interestedness effectively, but in tune with general principles and truths." In other words, the principal is concerned with immediate raw interests, while the agent is concerned with higher ideals. For example, the goals of a university as principal are to produce education and research at low cost, whereas the goals of the university as agent include having the maximum number of brilliant (expensive) professors and the maximum number of prestigious programs. The same tension manifests in individuals as classic psychological dualisms (e.g., short-term vs. long-term interests) By this duality, highly agentic features like opinions and attitudes can be decoupled from behaviors, actions, and decisions (Meyer and Jepperson, 2000).

Social agency, within this body of work, refers to the social standardization and scriptedness of agency, and to how agency dynamics permeate and shape social structure. In a society of social agents, each individual or organization acts in accordance with their socially prescribed and defined agency, which is akin to the ideals defining their social role. In general terms, "the actorhood of individuals, organizations, and national states [is] an elaborate system of social agency…" wherein actors routinely shift between agency for the self and otherhood for the generalized agency of the social system. Individuals share in the general social agency of the system, negotiating the bases for their own existence via the rules and definitions of the broader system. This general social agency can function as the capacity for collective agentic action (Meyer and Jepperson, 2000). This understanding of agency as an upholding of higher ideals, principles, and truths (and social agency as the collective version of this), often in conflict with baser self-interested principalhood, is so different from conceptions of agency and social agency in HRI as to be essentially completely disjoint concepts. As we will illustrate below, agency in HRI is not (to our knowledge) discussed in duality with the notion of a principal, and social agency is not understood as a collective version of individual agency.

In presenting the definitions in this section, we do not intend to suggest that other fields have reached some sort of internal consensus regarding social agency or perfect consistency in its usage. Like in HRI, there appears to be ongoing conversation and sometimes disagreement about social agency within many fields, though the HRI-specific branch of this conversation seems relatively nascent. For example, there are ongoing debates in anthropology about whether (social) agency is an essential property of individuals, or somehow exists only in the relationships between individuals. Likewise, there are differing opinions within and between social science research communities about whether nonhuman entities can have (social) agency Gardner (2016). Unfortunately, we cannot present all perspectives here, nor can we really present the full detail and nuance of some of the perspectives that we *have* presented. What we hope to have indicated is that definitions of social agency from other fields, though academically rigorous and undoubtedly useful within their respective domains, are, for various reasons, neither intended nor suitable for the unique role of social agency in HRI, and an HRI-specific definition is needed.

1.2 Theories of Social Agency in Human-Robot Interaction

A number of theories of Social Agency have been defined within the HRI community to address the unique perspective of our field. Many of these grew out of foundational work on Social Actors from Nass et al. (1994), which suggested that humans naturally perceive computers with certain characteristics (e.g., linguistic output) as social *actors*, despite knowing that computers do not possess feelings, "selves", or human motivations (Nass et al., 1994). This perception leads people to behave socially towards machines by, for example, applying social rules like politeness norms to them (Nass et al., 1994; Jackson et al., 2019). It is perhaps unsurprising that this human propensity to interact with and perceive computers in fundamentally social ways extends strongly to robots, which are often deliberately designed to be prosocial and anthropomorphised. While Nass et al.'s work establishing the theory that humans naturally view computers as social *actors* did not call computers "social agents" or refer to the "social agency" of computers, it nevertheless established that the human-computer relationship is fundamentally social, and laid the groundwork for much of the discussion of sociality and social agency in HRI today. In this section we will discuss four rigorously defined theories of Social Agency in HRI.

Nagao and Takeuchi

At around the same time that Nass and colleagues introduced their "Computers As Social Actors" (CASA) paradigm (Nass et al., 1994), Nagao and Takeuchi (1994) made one of the earliest references to computers as *social agents*. In describing their approach to social interaction between humans and computers, Nagao and Takeuchi argue that a computer is a social agent if it is both social and autonomous. These authors define socialness as multimodal communicative behavior between multiple individuals. Nagao and Takeuchi initially define autonomy as "[having] or [making] one's own laws," but later clarify that "an autonomous system has the ability to control itself and make its own decisions." We will see throughout this paper that sociality and autonomy remain central to our discussion of social agency today, but not necessarily as defined by these authors.

Nagao and Takeuchi also define a social agent as "any system that can do social interaction with humans," where a "social interaction" 1) involves more than two participants, 2) follows social rules like turn taking, 3) is situated and multimodal, and 4) is active (which might be better understood as mixed initiative). Some of these requirements, including at least the involvement of more than two participants and mixed initiativity, seem unique to this theory. Nagao and Takeuchi also differentiate their "social interactions" from problem solving interactions, though we believe, and see in the HRI literature, that task-oriented interactions can be social and take place among social agents.

Pollini

Pollini (2009) presents a theory that is less concerned with modality of interaction or type of robot embodiment, focusing instead on the role of human interactants in constructing a robot's social agency. For Pollini, robotic social agents are both physically and socially situated, with the ability to engage in complex, dynamic, and contingent exchanges. Social agency, then, arises as the outcome of interaction with (human) interlocutors, as "the ability to act and react in a goal-directed fashion, giving contingent feedback and predicting the behavior of others." We see the goal-directedness in this definition as loosely analogous to the notion of autonomy that is centered in other theories. In contrast to those theories, however, Pollini considers social agency as a dynamic and emergent phenomenon constructed collectively within a socially interacting group of autonomous actors, rather than as an individual attribute separately and innately belonging to the entities that comprise a social group. This presents a useful framing for understanding the social agency of multi-agent organizations like groups and teams. However, this multi-agent perspective prevents this definition from aligning with common references in HRI to the "social agency" of an individual robot. Nonetheless, some degree of autonomous behavior, interaction, perception, and contingent reaction must clearly remain central to our discussion of social agency.

Pollini also opines that "social agency is rooted in fantasy and imagination." It seems that humans' *attribution* of social agency may be tied to the development of imagination during childhood, leading Pollini to argue that people can "create temporary social agents" of almost anything with which they have significant contact, including toys like dolls, tools like axes, and places like the home. This leads them to the question "what happens when such 'entities-by-imagination' also show autonomous behavior and contingent reactions, and when they exist as social agents with their own initiative?" However, we argue that axes, dolls, and places actually *cannot* be social agents, at least not in the way that the typical HRI researcher means when they call a robot (or human) a social agent, since robots can conditionally take interactional behavior, which we believe is necessary for social agency.

Finally, Pollini argues that agency-specific cues embedded in robots (e.g., contingent behavior) are insufficient by themselves for creating social agency, and that social agency, rather, is negotiated between machines and their human interactants via a process of interpretation, attribution, and signification. This process involves interpreting a machine's behavior as meaningful and explicative, and then attributing social agency based on the signification of that behavior as meaningful, which may also involve attributing internal forces like intentions and motivations. This means that, through this process, things with simple behaviors like cars or moving shapes on a screen can end up being ascribed social agency. Again, however, we see a fundamental difference between these examples and social robots, which can actually deliberately manifest meaningful and explicative behaviors. We interpret this discussion as circling the distinction between "actual" and "perceived" social agency that we will discuss below.

Levin, Adams, Saylor, and Biswas

Though much of the HRI literature exploring the standalone concept of *agency* is beyond the scope of this work as it focuses on the agency of machines without centering notions of *sociality*, the theory of agency from Levin et al. (2013) is relevant here because it explores attributions of agency specifically during social human-robot interactions. Levin et al. argue that people's first impulse is to strongly differentiate the agency of humans and nonhumans, and that people only begin to equate the two with additional consideration (e.g., when prompted to do so by the robot defying initial expectations). They also describe how simple robot behavioral cues like the naturalness of movement or gaze can influence people's attribution of agency to robots, as well as states and traits of the human attributor, like loneliness. Like some previous theories, Levin et al. center goal-orientedness and intentionality in their account of agency. However, they include not only behavioral intentionality, which we saw in other theories (Pollini, 2009), but also intentionality in cognition. Their example of this cognitive intentionality is drawing ontological distinctions between types of objects based on their use rather than their perceptual features.

Ala

Finally, Alač (2016) presents a theory in which multimodal interaction, situatedness, and materiality are important to a robot's social agency, and justifies this theory with an observational study of a robot in a classroom. Alač frames robot agenthood as coexisting with the contrasting status of "thing," with agentic features entangled in an interplay with a robot's thing-like materiality. However, Alač moves away from discussing a robot's social nature as an intrinsic and categorical property that resides exclusively in the robot's physical body or programming, instead seeing robot sociality as enacted and emergent from how a robot is experienced and articulated in interactions. To Alač, the socially agentic facets of a robot are evident in the way it is treated by humans, focusing on proxemic and haptic interaction patterns and linguistic framing (e.g., gendering the robot) in group settings. Our work can augment ethnography-based theories like this one by exploring 1) the features of the *robot's* behavior that give rise to perceptions of social agency, 2) what concepts constitute such perceptions, and 3) exactly what such perceptions imply. In other words, we focus on what social agency *is*, rather than on human behaviors that indicate ascription thereof.

1.3 Notions of Social Agency in Human-Robot Interaction

While in the previous section we discussed rigorously defined theories of social agency, much of the HRI literature that engages with social agency does not actually connect with those theories. In this section, we will thus explore the ways in which HRI researchers casually refer to social agency without focusing on developing or defining a formal theoretical account of it. Our goals in doing so are to 1) illustrate that notions of social agents and agency are commonly applied within the HRI research community, 2) provide examples of *how* these terms are used, and demonstrate important qualitative differences among the entities to which these terms are applied, 3) show that the existing theories defined in the previous section do not capture the common parlance usage of "social agency" among HRI researchers, and 4) lay the groundwork for developing a theory that does accommodate these usages.

There are many papers that refer to robots as social agents without mentioning or dealing with *social agency* per se. The term social agent is widely applied to entities that are both embodied (Heerink et al., 2010; Lee et al., 2012; Luria et al., 2016; Westlund et al., 2016) and disembodied (Lee et al., 2006; Heerink et al., 2010); remote controlled by humans (Heerink et al., 2010; Lee et al., 2012; Westlund et al., 2016) and self-controlled (Heerink et al., 2010); task-oriented (Heerink et al., 2010; Lee et al., 2012) and purely social (Lee et al., 2006); anthropomorphic (Heerink et al., 2010; Lee et al., 2012), zoomorphic (Lee et al., 2006; Heerink et al., 2010; Westlund et al., 2016), and mechanomorphic (Heerink et al., 2010; Luria et al., 2016); mobile (Heerink et al., 2010; Lee et al., 2012) and immobile (Heerink et al., 2010; Luria et al., 2016); and able to communicate with language (Heerink et al., 2010; Lee et al., 2012) and unable to do so (Lee et al., 2006; Luria et al., 2016). Any theory of social agency for HRI, then, should either encompass this diversity of social agents or account for ostensible misattributions of social agency. However, the theories we have examined, which emphasize embodiment (Nagao and Takeuchi, 1994; Alač, 2016), language (Nagao and Takeuchi, 1994), and self-control or intentionality (Pollini, 2009; Levin et al., 2013), exclude usages that are apparently common in HRI research.

Of course, one could argue that casual references to robots as "social agents" are synonymous to references to robots as "social actors," and that such references do not actually have anything to do with the agentic nature of the robot. By this argument, the existing theoretical work on social agency in HRI would best be understood as investigating a completely separate topic from social agents. This reasoning, however, would result in a confusing state-of-affairs in which social agency is not a prerequisite for being a social agent, with the two topics unrelated except by the general connection to social interaction. We therefore assume that a social agent must be a thing with social agency, and that these two terms must be tightly and logically related. A clear conception of social agency is thus a prerequisite for the study of social agents. However, much of the work in HRI that concerns social agency does not focus on rigorously defining it. Indeed, some of these studies do not explicitly provide their definition of social agency at all.

An illustrative example of a casually referenced "social agent" is the "Snackbot" developed by Lee et al. (2012). The anthropomorphic Snackbot had real interactions with many humans over the course of multiple months as a snack delivery robot. The robot's movement was self-controlled, but a human teleoperator hand-selected its delivery destinations. The human operator also remotely controlled the robot's head and mouth movements and the robot's speech, by selecting from a number of pre-made scripts, both purely social and task-oriented. We will refer back to this example in **Section 2**.

In their investigation of how cheating affects perceptions of social agency, Ullman et al. (2014) used perceptions of trustworthiness, intelligence, and intentionality as indicators of perceptions of social agency in an anthropomorphic robot. Using intentionality as a proxy for social agency aligns directly with several of the theories that we described in **Section 1.2** (Pollini, 2009; Levin et al., 2013). Intelligence and trustworthiness, however, seem less closely related to social agency, and trustworthiness is explicitly not an aspect of social agency in theories that discuss competition and uncooperative behavior as inherently social actions (Castelfranchi, 1998).

Baxter et al. (2014) also study attributions of social agency to robots without explicitly defining the term, and measure it via a different proxy: human gaze behavior. This proxy does not obviously align with any of the theories of social agency discussed above. Although it is possible that gaze could be a good proxy for some definition of social agency (or the ascription thereof), further empirical work would be needed to establish that relationship.

Straub (2016) adopt yet another definition of social agency in their investigation of the effects of social presence and interaction on social agency ascription. In their study, social agents are characterized as "having an 'excentric positionality,' equipped with a) an ability to distinguish themselves, their perceptions as well as their actions from environmental conditions (embodied agency), b) the ability to determine their actions and perceptions as self-generated, c) having the ability to define and relate to other agents equipped with the same features of a) and b), along with d) defining their relationship to other agents through reciprocal expectations toward each other ('excentric positioned' alter ego)."

This definition, particularly part b, is somewhat ambiguous. One interpretation is that the robot simply needs to distinguish its own actions from the actions of others, and know that it is the cause for the effects of its actions; if the robot moves its arm into a cup, then it is the source for both the movement of the arm and the movement of the cup. However, this seems more like the robot knowing that its actions' effects are self-generated and that it was the one that acted, rather than viewing the choice to act or the genesis of the action itself as self-generated. Another interpretation, which is similar to some of the definitions of social agency discussed in **Section 1.1**, is that seeing an action as self-generated requires the robot to understand its choice to act, perceive that choice as its own, and believe that it could have acted differently. This definition appears to require some form of consciousness or experience of free will, and is thus not well-suited to HRI. Straub uses human behavioral proxies, like eye contact, mimicry, smiles, and utterances, to measure ascriptions of social agency to robots (with more of these behaviors indicating more ascribed social agency), but such behavioral proxies do not measure all components of their definition.

Ghazali et al. (2019) study the effects of certain social cues (emotional intonation of voice, facial expression, and head movement) on ascriptions of social agency. Professedly inspired by research in educational psychology described above (Atkinson et al., 2005), they define social agency as "the degree to which a social agent is perceived as being capable of social behavior that resembles human-human interaction," and then measure it by collecting participant assessments of the extent to which the robot was "real" and "like a living creature." Roubroeks et al. (2011) use the exact same definition of social agency as Ghazali et al. (2019) in their investigation of psychological reactance to robots' advice or requests, but operationalize it differently. Although they did not attempt to measure social agency, they did seek to manipulate it by varying robot presentation, presenting a robot's advice as either text alone, text next to a picture of the robot, or a video of the robot saying the advice.

This definition seems problematically circular in that it defines social agency by the degree to which a social agent does something, without defining what it means to be a social agent. We also argue that Ghazali et al.'s chosen measures do not clearly align with the formal definitions of social agency proposed above, nor with Ghazali et al.'s stated definition. Moreover, this conceptualization excludes a large number of robots that the HRI literature calls social agents, and focuses on factors that many theories de-emphasize (e.g., livingness and human likeness). This example in particular shows that disparate definitions of social agency currently exist in the HRI literature, leading to confusion when authors underspecify or neglect to specify a definition.

Other work from Ghazali et al. (2018) on the relationship between social cues and psychological reactance centers the concepts of "social agent" and "social agency" explicitly, using the terms over 100 times in reference to robots and computers. However, the authors do not expressly provide any definition for those terms, despite ostensibly manipulating social agency in an experiment. Implicitly, the authors appear to follow their definition described above, with more humanlike superficial behavior (e.g., head/eye movement and emotional voice intonation) being considered more socially agentic, while the semantic content and illocutionary force of all utterances was kept constant across social agency conditions. However, Ghazali et al. (2018) also seem to consider the capacity to threaten others' autonomy as a critical feature of social agency, since they measure perceived threat to autonomy as a manipulation check on social agency (though the social agency manipulation did not significantly impact perceived threat to autonomy). This choice was not extensively justified. As discussed in **Section 2.2**, perceived threat to autonomy is strongly related to (negative) face threat, which we view as important to social agency. However, as we will discuss, the capacity to threaten face is far broader than the capacity to threaten autonomy as measured by Ghazali et al. (2018).

To summarize, we have discussed several conflicting theories and usages of social agency in HRI, which, to varying extents: a) exclude common uses of the term "social agency" by being too restrictive, b) include objects that nearly all researchers would agree are neither social nor agentic, c) focus on factors that do not seem relevant to social agency in most pertinent HRI work, or d) conflate other concepts (like livingness or human-likeness) with social agency as it seems commonly understood. In addition, we have shown examples of the diversity of uses of the term "social agency" in the HRI research literature. We now contribute our own theory of social agency, with the specific intention of accommodating the HRI research community's existing notions of social agency.

2 A THEORY OF SOCIAL AGENCY FOR HUMAN-ROBOT INTERACTION

In this section, we propose a formal theory of social agency for HRI to address the challenges and limitations discussed in the previous sections. Our key arguments are: 1) social agency may be best understood through parallels to moral agency; 2) considering various levels of abstraction (LoAs) is critical for theorizing about any kind of agency; 3) a social agent can be understood as something with agency that is capable of social action; 4) social action is grounded in face; and 5) social and moral agency are related yet independent.

To best understand social agency, we draw parallels to recent work on moral agency. Not only are the concepts centered in theories of social agency discussed in **Section 1.2** (e.g., autonomy, contingent behavior, and intentionality) also centered in many theories of moral agency, but the moral agency of robots and other artificial actors has also received a more rigorous treatment than social agency in the HRI literature. The moral agency literature thus represents a valuable resource for constructing a parallel theory of social agency. Furthermore, the two concepts of moral and social agency are inexorably linked, representing the two halves of interactional agency. They provide congruent relationships to (and means of understanding) moral/social norms and are key to our most foundational understandings of interaction. Given these similarities and connections, parallel understandings of the two concepts are not only intuitive but necessary, and we see no reason to attempt to define moral and social agency completely separately. For our purposes, we will leverage the moral agency theory of Floridi and Sanders (2004), but note that, as with social agency, there is not yet consensus among scholars as to a single canonical definition of moral agency, prompting ongoing debate (Johnson and Miller, 2008).

2.1 Agency and Levels of Abstraction

Because of historical difficulties in defining necessary and sufficient conditions for agenthood that are absolute and context-independent, Floridi and Sanders (2004) take analysis of *levels of abstraction* (LoAs) (Floridi, 2008) as a precondition for analysis of agenthood. A LoA consists of a collection of observables, each with a well-defined set of possible values or outcomes. An entity may be described at a range of LoAs. For a social robot, the observables defining an average user's LoA might only include the robot's behavior and other external attributes, like robot morphology and voice. In contrast, the robot developer's LoA would likely also include information internal to the robot, such as the mechanisms by which it perceives the world, represents knowledge, and selects actions. Critically, a LoA must be specified before certain properties of an entity, like agency, can be sensibly discussed, as a failure to specify a LoA invites inconsistencies and disagreements stemming not from differing conceptions of agency but from unspoken differences in LoA.

The "right" LoA for discussing and defining moral agency must accommodate the general consensus that humans are moral agents. Floridi and Sanders (2004) propose a LoA with observables for the following three criteria: interactivity (the agent and its environment can act upon each other), autonomy (the agent can change its state without direct response to interaction), and adaptability (the agent's interactions can change its state transition rules; the agent can "learn" from interaction, though this could be as simple as a thermostat being set to a new temperature at a certain LoA). For the sake of simplicity, we will consider LoAs consisting only of observations that a typical human could make over a relatively short temporal window. These observables encompass some concepts that were important to the theories discussed in **Section 1.2** (e.g., autonomy and contingent behavior), and exclude others (e.g., teleological variables like intentionality or goal-directedness), which we discuss more below. We also consider a criterion that was *not* included in many theories for social agency, namely adaptability.

At the user's LoA, wherein the deterministic algorithms behind a robot's behavior are unobservable, the robot is interactive, autonomous, and adaptable, and therefore is an agent. However, at the robot developer's LoA [or what Floridi and Sanders (2004) call the "system LoA"], which includes an awareness of the algorithms determining the robot's behavior, the robot loses the attribute of adaptability and is therefore not an agent. These two LoAs will be important throughout the rest of this paper.

We argue that the distinction between these two LoAs (the user's and the developer's) explains why some scholars have suggested conceptualizing and measuring "*perceived* moral agency" in machines as distinct from moral agency itself. This notion of perceived moral agency would ostensibly capture "human attribution of the status of a machine's agency and/or morality (independent of whether it actually has agency or morality)" (Banks, 2019), and these authors could easily define "perceived social agency" the same way.

Much of the impetus for defining these new concepts seems to be a desire to avoid the varied and conflicting definitions for agency (and the social and moral variants thereof). Typically within HRI, researchers are primarily concerned with how their robots are *perceived* by human interactants (the user's LoA), and how those interactants might ascribe social agency to those robots. In that sense, perceived social agency as a concept seems like a good way to allow researchers to focus on what they really care about without getting mired in discussions of their robot's "actual" agency, though it can still leave exactly what is perceived as (socially) agentic underspecified.

However, as we saw in **Section 1**, authors seldom refer to perceived social agency (particularly since we just defined it as parallel to perceived moral agency, which also does not seem to have caught on), but rather use the unqualified term "social agency". Thus, rather than attempting to enforce a change in terminology, we propose that "perceived moral/social agency" should be understood as moral/social agency at the robot user's LoA, and "actual" moral/social agency is the corresponding notion at the developer's LoA. To illustrate, consider the SnackBot (Lee et al., 2012) described in **Section 1.3**. This robot was largely remotely controlled by a human, but, at the snack orderer's (user's) LoA it is a social agent. At the developer's LoA, the robot is not an agent, but the system in aggregate might be considered socially agentic since one of its constituent parts, the human, is a social agent in and of itself.

If SnackBot could manifest the same behavior without human input, it would still not be agentic at the developer's LoA insofar as its behavior is the direct result of deterministic algorithms that only act on its state. However, it does intuitively *seem* more agentic, prompting us to consider another useful LoA: one where we are aware of the general distributed system that controls a robot (in terms of software cognitive architectural components, hardware components like cloud computing, and human teleoperators), but not aware of the inner workings of each constituent part of that system. At this LoA, which we call the "architecture LoA", a robot that does its computation internally might be agentic, but a robot that is remote controlled by either a person or another machine could not be an agent in and of itself. Hundreds of different LoAs could be constructed with various degrees of detail regarding how a robot works, but this is largely not constructive if humans are unlikely to ever view the robot from those LoAs. However, we believe that the architecture LoA is realistic for many potential robot interactants, particularly those that might own their own personal robots, or participants in laboratory HRI studies after the experimental debriefing.

At first glance, it would be easy to draw some parallels between our three main LoAs (developer's, architecture, and user's) and Dennett's three stances from which to view an entity's behavior in terms of mental properties (physical, design, and intentional) (Dennett, 1978). The user's LoA in particular bears loose resemblance to Dennett's intentional stance because the user is aware only of the robot's externally observable behaviors, and may rationalize them by projecting internal states onto the robot. Likewise, our architecture LoA is explicitly concerned with the parts comprising a robot's distributed system and the broad purpose of each constituent part, like the design stance, though it is not necessarily concerned with the purpose of the robot itself as a whole. However, several key distinctions separate our three LoAs form Dennett's three stances. Most obviously, the developer's LoA is unlike Dennett's physical stance in that it is concerned with the algorithms producing the robot's behavior but not the specifics of their implementation nor the hardware executing them.

More broadly, the three LoAs we have presented generally represent three of the *sets of information* that real people are most likely to have regarding robots during HRI, but there is no reason for this set of LoAs to be considered exhaustive, and no reason why our analysis of social agency cannot also apply to any other LoA from which a person views a robot. In contrast, more rigidly tripartite approaches to epistemological levelism, like Dennett's, though readily formalized in terms of LoAs, contain an implicit ontological commitment and corresponding presupposed epistemological commitment because they privilege explanations over observable information (Floridi, 2008). That is not to say that such approaches to multi-layered analysis are not interesting and illustrative to HRI. For example, many researchers have explored whether humans naturally adopt the intentional stance towards robots and other artificial entities like they do towards other humans (Thellman et al., 2017; Marchesi et al., 2019; Perez-Osorio and Wykowska, 2019; Schellen and Wykowska, 2019; Thellman and Ziemke, 2019). However, it seems intuitive that robot developers versus users might take the intentional stance towards robots to different extents and under different conditions, so we posit that a specification of LoA is helpful in considering Dennett's stances and other attitudinal stances in HRI in much the same way that it is to our discussion of social agency, rather than Dennett's stances being homeomorphic to the three LoAs most salient here.

Most current cognitive architectures are precluded from agency at the developer's LoA because any learning is typically a matter of updating the robot's state by the deterministic rules of its code, rather than an actual update to the rules for transitioning between states (Floridi and Sanders, 2004). This includes black-box systems, like deep neural networks, because their lack of interpretability comes from an inability to fully understand how the state results in behavior, not from actual adaptability. However, we accept that humans have adaptability, and see no theoretical reason why the same level of adaptability could not be implemented in future artificial agents. Of course, particularly within the theory of causal determinism, there exists an LoA wherein humans do not have agency if all human behavior is rooted in the physical and chemical reactions of molecules in the brain (a "physical" LoA *a la* Dennett). Regardless of the veracity of this deterministic point of view, it seems clear that no LoA precluding agency from existing in the universe as we know it is a useful LoA at which to discuss agency in HRI.

We adopt the above notion of LoA and criteria for agenthood from Floridi and Sanders (2004) for our theory of social agency for several reasons. First, different LoAs help us to account for different understandings of social agency in the HRI literature, as we saw in our discussion of "actual" versus "perceived" social agency. Second, we can explicitly avoid conflating moral/social *agency* with moral/social *responsibility* (i.e., worthiness of blame or praise), which is another discussion beyond the scope of this paper. Third, avoiding internal variables like intentionality, goal-directness, and free-will guarantees that our analysis is based only on what is observable and not on psychological speculation, since a typical robot user cannot observe these attributes in the internal code or cognitive processes of their robot; we thus prefer a phenomenological approach.

Having established an understanding of agency, we now need to define some notion of sociality congruent to Floridi and Sanders's notion of morality. However, we first want to point out that our justification for avoiding unobservable factors in defining and assessing (moral/social) agency parallels a similar argument from proponents of ethical behaviorism in defining and assessing the moral status of robots. Ethical behaviorism is an application of methodological behaviorism (as opposed to ontological behaviorism) to the ethical domain, which holds that a sufficient reason for believing that we have duties and responsibilities toward other entities (or that they have rights against us) can be found in their observable relations and reactions to their environment and ourselves. In other words, robots have significant moral status if they are roughly performatively equivalent to other entities that have significant moral status, and whatever is going on unobservably "on the inside" does not matter. This is not to say that unobservable qualia do not exist, nor do we deny that such qualia may be the ultimate metaphysical ground for moral status. However, the ability to ascertain the existence of these unobservable properties ultimately depends on some inference from a set of observable representations, so a behaviorist's point of view is necessary to respect our epistemic limits (Danaher, 2020). We agree with this reasoning. Our definition of social agency could be framed as a form of "social behaviorism" that specifies the behavioral patterns that epistemically ground social agency and, by considering LoAs, is sensitive to the behaviors that are actually observed, rather than the set of behaviors that are, in principle, observable.

Of course, avoiding attributes like intentionality or goal directedness in our definitions in favor of a behaviorist approach does not completely free us from needing to rely on some form of inference. At a minimum, making observations from sensory input requires the inference or faith that one's sensory inputs correspond to some external reality. Likewise, our interactivity criterion for agency requires some causal inference or counterfactual reasoning. For example, concluding that a robot can be acted on by the environment requires the counterfactual inference that the robot's "response" to a stimulus would not have occurred absent that stimulus. Unfortunately, requiring some inference is unavoidable. In light of this, one could argue that it is equally reasonable and necessary to infer intention and goal directedness from behavior. For example, pulling on a door handle might signal an intent to open the door with the goal of getting into the building, even though the same behavior could also signal mindless programming to tug on handles without representing goals or having intentions. We argue that the sensory and causal inferences required by our framework are lesser epistemological leaps and more necessary and common (and therefore more justifiable) than inferences about other agent's mental states like intentionality and goals. We also emphasize that goals and intentions are apparently not important to social agency at the developer's LoA, since we saw many robots referred to as social agents by their developers in **Section 1.3** that did not internally represent goals or intentions, and their developers would have known that.

2.2 Social Action Grounded in Face

We now move on to developing a notion of sociality congruent to Floridi and Sanders's notion of morality. For Floridi and Sanders (2004), any agent that can take moral action on another entity (e.g., do good or evil; cause harm or benefit) is a moral agent. Any entity that can be the recipient of moral action (e.g., be harmed or benefited) is a moral patient. Most agents (e.g., people) are both moral agents and moral patients, though research has indicated an inverse relationship between perceptions of moral agency and moral patiency (e.g., neurodivergent adults are perceived more as moral patients and less as moral agents than neurotypical adults) (Gray and Wegner, 2009).

Just as a *moral* agent is any agentic source of moral action, we can define a *social* agent as any agentic source of social action. We ground our definition of social action in the politeness theoretic concept of "face" (Brown and Levinson, 1987). Face, which consists of positive face and negative face, is the public self-concept (meaning self-concept existing in others) that all members of society want to preserve and enhance for themselves. Negative face is defined as an agent's claim to freedom of action and freedom from imposition. Positive face consists of an agent's self-image and wants, and the desire that these be approved of by others. A discourse act that damages or threatens either of these components of face for the addressee or the speaker is a face threatening act. Alongside the level of imposition in the act itself, the degree of face threat in a face threatening act depends on the disparity in power and the social distance between the interactants. Various linguistic politeness strategies exist to decrease face threat when threatening face is unavoidable or desirable. Conversely, a face affirming act is one that reinforces or bolsters face for the addressee or speaker (though our focus will be on the addressee). We define social action as any action that threatens or affirms the addressee's face. So, affirming and threatening face are social analogs to doing moral good and harm respectively. In contexts where it is helpful, this definition also allows us to refer to robots with different capacities to affect face as having different degrees of social agency, rather than viewing social agency as a strictly binary attribute. We also propose that the term "social *actor*" can refer to interactive entities capable of social action, but lacking the other criteria for agency (autonomy and/or adaptability).

Some scholars have opined that it is common to view social agents as equivalent to "communicating agents" (Castelfranchi, 1998), and thus might simply say that any communicative action is a social action. Though the ability to nontrivially communicate implies the capacity to threaten face, we choose to base our definition of social action directly on face because it allows for a more intuitive parallel to moral agency without excluding any meaningful communicative actions. The vast majority of communicative actions that an agent can perform have the capacity to impact face. Just in terms of face threat, any kind of request, reminder, warning, advice, offer, commitment, compliment, or expression of negative emotion threatens the addressee's negative face, and any criticism, rebuke, insult, disagreement, irreverence, boasting, non-cooperation, or raising of divisive topics threatens the addressee's positive face (Brown and Levinson, 1987). A single speech act can carry several elements that affect face in different ways, and even the mere act of purposefully addressing someone is slightly affirming of their positive face by acknowledging them as worth addressing, and slightly threatening of their negative face by imposing on their time. Indeed, it is difficult to think of a meaningful communicative action that would have no impact on face.

Another reason to ground social action in face is because face is more concrete and computationalizable than some other options (e.g., induced perceptions of human likeness or influence on emotional state), while still being broad enough to encompass the whole set of actions that we would intuitively consider to be social. There exist various parameterizations or pseudo-quantifications of face threat/affirmation, including Brown and Levinson's own formula which presents the weight of a face threatening act (W) as the sum: $W = D(S, H) + P(H, S) + R$ where $D(S, H)$ is the social distance between the speaker (S) and hearer (H), $P(H, S)$ quantifies the power that H has over S, and R represents the culturally and situationally defined level of imposition that the face threatening act entails. For negative face threatening acts, R includes the expenditure of time and resources. For positive face threatening acts, R is harder to determine, but it is given by the discrepancy between H's own desired self-image and that presented in the face threatening act. Individual roles, obligations, preferences, and other idiosyncrasies are subsumed into R. Of course, the constituent parts of this equation cannot be precisely quantified in any canonical way (nor can, for example, influence on behavioral or emotional status). We do not view this as a weakness because we would not expect to precisely quantify the magnitude of socialness in an action. Humans cannot precisely answer questions like "How social is it to hug your grandmother?" or "Which is more social, asking a stranger for the time or tipping your waitress?". However, this equation nonetheless illustrates some of the concrete underpinnings of face and shows how face connects to concepts like relational power, interpersonal relationships, material dependence, cultural mores, etc.

Robots are valid sources of social action under this face-based definition. Typical task-oriented paradigms of HRI involve robots either accepting or rejecting human requests (which either affirms or threatens both positive and negative face), or making requests of humans (which threatens negative face). Even simply informing human teammates about the environment threatens negative face by implying that the humans ought to act based on the new information. Less task-oriented cases, like companionship robots for the elderly (Heerink et al., 2010), also require face affecting social actions, though these may tend to be more face affirming than in task-based interaction. Again taking the SnackBot Lee et al. (2012) as an example, bringing someone a requested snack is face affirming, and so are dialogue behaviors like complimenting snack choice or apologizing for delays. The SnackBot's dialogue behavior of asking people to move out of the way is face threatening. Research examining how robots influence human face and how humans react to robotic face threatening actions is ongoing (Jackson et al., 2019; Jackson et al., 2020).

In comparison to our definition, Castelfranchi (1998) define an action as either social or nonsocial depending on its purposive effects and the mind of the actor. Their social actions must be *goal-oriented* and motivated by *beliefs* about predicted effects in relation to some goal. Their social actions are mainly based on some exercise of power, to attempt to influence the behavior of other agents by changing their minds. They specifically say that social action cannot be a behavioral notion based solely on external description. This definition is not well-suited to our purposes because these internal underpinnings are unknowable to a typical robot user, and thus preclude the user from viewing a robot as a social agent. We saw similar reasoning in our decision to exclude goal-orientedness as a prerequisite for agency. Even if a user chooses to adopt an intentional stance (see Dennett, 1978) toward a robot and infer goals motivating its behavior, this does not imply that the robot actually has an internal representation of a goal or of the intended effects of its actions; the person's intentional stance would only allow them to take social action towards the robot, not vice versa. Given the popular perception of robots as social and the academic tendency to call them social agents, we do not want a definition of social action that cannot apply to robot action or that relies on factors that cannot be observed from a user's LoA. Furthermore, Castelfranchi's definition excludes, for example, end-to-end deep neural dialogue systems that may not explicitly represent goals, beliefs, causality, or interactants as potential sources of social action, but whose actions can clearly come across as social and carry all the corresponding externalities. Our face-based definition does not have these limitations.

To be clear, our decision to define social action via face is not an arbitrary design choice, but rather a result of face's integral role in all social interaction. We believe that an action's relationship to face is, unavoidably and fundamentally, what determines whether that action is social because face is what creates the experience of having social needs/desires in humans. It follows that, for robots, the appearance or attribution of face, or some relationship to others' face, is what allows them to be social actors. Any action that affects face is necessarily social, and any action that does not is necessarily asocial. This aligns well with widespread intuitions about sociality and common parlance use of the term.

2.3 Social Patiency as Having Face

Any social action must have a recipient whose face is affected. If social agency is an agent's capacity to be a source of social action (to affirm or threaten face), then the corresponding notion of social *patiency* is the capacity to have one's face threatened or affirmed (i.e., having face). This is similar to the notion of moral patiency as the capacity to be benefited or harmed by moral action. Clearly, conscious humans are simultaneously moral and social agents and patients at any reasonable LoA. However, neither moral nor social patiency at any given LoA strictly requires moral or social agency at the same LoA, which leads us to the question of whether our robotic moral/social agents in HRI are also moral/social patients.

It seems clear that, at a reasonable LoA for a human interactant, it is possible to harm a robot, making the robot a moral patient. This is especially clear for robots capable of affective displays of protest and distress (Briggs and Scheutz, 2014). Indeed people deliberately abuse robots with surprising frequency (Nomura et al., 2015). However, at a deeper LoA, we know that current robots cannot feel pain (or pleasure), have no true internal emotional response to harm like fear, and lack the will towards self preservation inherent in most lifeforms. Thus, at this deeper LoA the robot is not a moral patient.

Likewise, a robot's social patiency depends on the LoA considered. It is feasible to program a robot to manifest behaviors indicating face wants, like responding negatively to insults and positively to praise, in which case it would be a social patient at the user's LoA. However, at the developer's LoA, the robot still has no face.

2.4 Social and Moral Agencies as Independent

We now discuss the extent to which social agency and moral agency can manifest in machines independent of one another. We believe that some machines, including some robots, are largely perceived as asocial moral agents, while others are seen as amoral social agents. Although, for the most part, social robots do not fall in either of these groups, we believe that they are worth presenting as points of reference for understanding the special moral and social niche occupied by language capable robots. We continue to consider these technologies from the user's LoA.

Some artificial agents are popularly ascribed some form of moral agency without behaving socially or even possessing the capacity for communication outside of a narrow task-based scope. We call such agents "asocial moral agents", and use autonomous motor vehicles as the quintessential example. If we include the likely possibility that autonomous vehicles will learn and change their behavior in response to changing road conditions or passenger preferences, they are agentic at the passenger's LoA by being interactive, autonomous, and adaptive.

In terms of moral action, while autonomous motor vehicles are obligated to conform to the legal rules of the road, they are also expected to engage in extralegal moral decision making and moral reasoning. Myriad articles, both in popular culture and in academia, contemplate whether and how autonomous cars should make decisions based on moral principles (e.g., Bonnefon et al., 2016). Questions like "in an accident, should the car hit a school bus to save its own passenger's life? Or should it hit the barrier and kill its passenger to save the school children?" have taken hold of popular imagination and proliferated wildly. Regardless of the actual usefulness of such questions (cf. Himmelreich, 2018), it is clear that autonomous cars are being ascribed moral agency.

We can also consider whether autonomous vehicles might be capable of social action. For example, using a turn signal is clearly communicative, but it is also legally mandated; an autonomous vehicle would signal an impending turn regardless of whether any other driver was present to see the turn signal. Given the legal motivation behind the turn signal and the fact that it has no specific intended addressee, we view it as the rare communicative act with no (or negligible) impact to face. Indeed, any communication via turn signal would be considered incidental to law-following by the typical driver. Other driving behavior can also be communicative; though we do not expect autonomous vehicles to engage in tailgating or road rage, we could imagine that they might change the norms governing human driving behavior by modeling those norms themselves. For example, if all autonomous vehicles on the road adopt a uniform following distance, this behavior might influence human drivers sharing the road to do the same. However, this potential normative influence is distinct from that of social robots in that it is passive, incidental, unintentional, and not principally communicative, and therefore not face-relevant.

In other cases, depending on behavior, robots could be perceived as amoral social agents. Social robots that do not have the ability to act on their environment in any meaningful extra-communicative capacity may be physically unable (or barely able) to produce moral action. As an example, consider MIT's Kismet robot, which is expressive, (non-linguistically) communicative, and social, but largely helpless and incapable of acting extra-communicatively. Many social actions are available to Kismet. For example, making a happy expression/noise when a person enters the room is face affirming, and a disgusted expression face threatening. Given the right behaviors, Kismet could also meet our prerequisites for agency and be an amoral social agent.

When moral and social agency are both present, as is the case for most social robots at the user's LoA, their combination gives rise to interesting phenomena. Social robots can occupy a unique sociotechnical niche: part technological tool, part agentic community member. This status allows robots to play an active role in shaping the community norms that inform human morality, which behavioral ethics has shown to be dynamic and malleable (Gino, 2015). And while robots are not the only technology to play a role in shaping human norms (Verbeek, 2011), we believe their social agency grants them uniquely powerful normative influence. For example, robots have been shown to hold measurable persuasive capacity over humans, both via explicit and implicit persuasion (Briggs and Scheutz, 2014; Kennedy et al., 2014), and even to weaken human (application of) moral norms via simple question asking behavior (Jackson and Williams, 2019).

Language capable robots are unique among technologies not only in the strength of their potential moral influence, but also in their ability to take an active and purposeful role in shaping human moral norms (or human application of moral norms) as social agents. However, this capability is a double-edged sword. On the one hand, robots of the future could productively influence the human moral ecosystem by reinforcing desirable norms and dissuading norm violations. On the other hand, today's imperfect moral reasoning and natural language dialogue systems open the door for robots to inadvertently and detrimentally impact the human moral ecosystem through reasoning errors, miscommunications, and unintended implicatures. It is thus crucial to ensure moral communication and proper communication of moral reasoning from robots, especially in morally consequential contexts. The power to transfer or alter norms comes with the responsibility to do so in a morally sensitive manner.

3 REVISITING RELATED WORK

Revisiting the theories of social agency from **Section 1.2**, we see that our definition is more inclusive than that of Nagao and Takeuchi (1994) and Alač (2016) in that we demphasize the robot's embodiment and materiality to account for purely digital potential social agents that we see in HRI research (Lee et al., 2006; Heerink et al., 2010), and do away with the teleological and internal considerations (e.g., goal-orientedness and intentionality) that would not be knowable to the typical robot user (cp. Pollini, 2009; Levin et al., 2013). On the other hand, our work is more restrictive than Pollini (2009) because we exclude "entities by imagination" as potential social agents, and specify that there are several behavioral traits necessary for social agency. This approach balances the more human-ascription-centered and more robot-trait-centered conceptualizations of social agency. Our theory acknowledges the human role in determining social agency by centering human face and the human's LoA, without reducing social agency to the mere ascription thereof. At the same time, we concretely describe the robot traits necessary for social agency at a given LoA.

Revisiting the studies from **Section 1.3**, which referenced social agents and social agency without principally focusing on defining those concepts, we see that our definition can encompass the wide diversity of potential social agents in HRI. Particularly at the user's LoA, robots can be social agents regardless of embodiment, teleoperation, task-orientedness, morphology, mobility, or linguistic capacity. However, some of the robots we reviewed would actually be excluded by our definition at the user's LoA by failing to meet behavioral prerequisites, particularly by lacking indications of adaptability (e.g., Lee et al., 2006; Heerink et al., 2010; Roubroeks et al., 2011). Interestingly, robots with a human teleoperator, like the SnackBot (Lee et al., 2012) might be *more* likely to be socially agentic at the user's LoA than those with simpler self-controlled behavior.

Finally, we stress that our theory complements (rather than competes with) much of the previous work we discussed. For example, some of the proxemic and haptic human behavior that Alač (2016) observed in their ethnographic study, like the choice to touch a robot's forearm rather than other body parts, might be understood within our theory as stemming from attributions of social *patiency* to the robot, rather than social agency. Likewise, our conception of social agency may well be tied to, for example, psychological reactance (Roubroeks et al., 2011) or trust (Ullman et al., 2014).

4 CONCLUDING REMARKS

We have presented a theory of social agency wherein a social agent (a thing with social agency) is any *agent* capable of *social action* at the *LoA* being considered. A LoA is a set of observables, and the LoAs most relevant to our discussion have been the robot user's, the developer's (or system LoA), and, to a lesser extent, the architecture LoA. *Agency* at any given LoA is determined by three criteria which we defined concretely above: interactivity, autonomy, and adaptability. We have defined *social action* as any action that threatens or affirms the addressee's face, and refer to the addressee in this scenario as a social patient. More specifically, *social patiency* is the capacity to be the recipient of social action, i.e., having face. These definitions came from parallel concepts in the philosophy of *moral* agency (Floridi and Sanders, 2004). We motivated our theory of social agency by presenting a sample of the inconsistent, underspecified, and problematic theories and usages of social agency in the HRI literature.

Based on our theory, we have several recommendations for the HRI community. We recognize a tendency to casually use the word "agent" to refer to anything with any behavior, and to correspondingly use "social agent" to simply mean "social thing." A summary of the concepts that are central to our theory can be found in **Table 1**. We encourage authors to consider either switching to the broader term "social actor" as defined above, or to briefly specify that they are using the term "social agent" informally and do not intend to imply social agency in any rigorous sense. We further recommend that any paper dealing with social agency be specific in selecting a suitable definition (such as the one presented in this work) and LoA.

It will be important for future studies to develop, refine, and validate measurements of social (and moral) agency. There exists early work on developing a survey to measure "perceived moral agency" for HRI (Banks, 2019), however some questions seem to conflate moral *goodness* with moral *agency*, and, despite measuring facets of autonomy and moral *cognition*, the survey does not measure the capacity for taking moral *action*. Some of the proxies that we saw used for social agency in **Section 1.3**, like human-likeness, realness, and livingness (Ghazali et al., 2019) do not match our new conceptualization of social agency. Others, like gaze (Baxter et al., 2014), could be promising but have yet to be validated with our theory (or, to our knowledge, any particular theory) of social

TABLE 1 | Summary of terms that are important to our concept of social agency.

Term	Definition
Level of Abstraction (LoA)	A collection of observables describing an entity (Floridi and Sanders, 2004; Floridi, 2008). A user's LoA for a robot includes movement, speech, morphology, etc., while the developer's LoA also includes the algorithms controlling the robot
Agent	Anything possessing the three criteria of interactivity, autonomy, and adaptability
Interactivity	The capacity to act on the environment and to be acted upon by the environment (Floridi and Sanders, 2004)
Autonomy	The capacity to change state without direct response to interaction (Floridi and Sanders, 2004)
Adaptability	The capacity for interaction to change the system's state transition rules. The capacity to "learn" from interaction (Floridi and Sanders, 2004)
Social agent	Anything capable of taking social action at the LoA under consideration
Social action	Any act that threatens or affirms an other's face. Analogous to moral action (doing harm/good to an other)
Social patient	Anything that can be a recipient of social action, i.e., anything with face
Face	The public self-concept (meaning self-concept existing in others) that all members of society want to preserve and enhance for themselves Negative face: an individual's claim to freedom of action and freedom from imposition Positive face: an individual's self-image and wants, and the desire that these be approved of by others (Brown and Levinson, 1987)

agency in mind. Validated metrics would facilitate experimental work motivated by our theory.

For example, future work designed to evaluate and further concretize our theory could empirically verify whether changing the LoA at which somebody is viewing a robot causes a corresponding change to their assessment of that robot as a (social) agent. The results could either strengthen the argument that the LoA is a critical prerequisite for the discussion of agency, or indicate that colloquial conceptions of agency do not account for LoA, despite its importance in rigorous academic discussions. Another avenue for this type of work would be to manipulate the magnitude of face threat/affirmation that a social robot is capable of and examine how that manipulation effects perceptions of the robot as a social agent. This experiment would specifically target our definition of social action as grounded in face.

Measures of social agency would also allow us to examine its relationship with persuasion and trust. On the one hand, we could imagine that decreasing a robot's social agency (by lowering its propensity to affect face) could increase its persuasive capacity if people are more amenable to persuasion when their face is not threatened. On the other hand, increasing a robot's social agency might increase its persuasive capacity if people are more likely to trust a more human-like robot.

Furthermore, it will be important to probe for causal relationships between ascriptions of social agency and ascriptions of moral responsibility and competence in robots. In human children, development of increased capacity for social action is typically correlated with development of other facets of intelligence and skills, including moral reasoning. However, this correlation does not necessarily exist for robots, since a robot could be socially agentic and competent, with a wide range of possible social actions, and still have no moral reasoning capacity. If robot social agency, or social behavior in general, leads interactants to assumptions of moral competence or overall intelligence (as it likely would in humans), this could lead to dangerous overtrust in robot teammates in morally consequential contexts that they are not equipped to handle. Thus, giving a robot linguistic/social competence would also necessitate giving the robot a corresponding degree of moral competence.

Finally, though there is evidence for an ontological distinction between humans and robots (Kahn et al., 2011), it is not yet clear where differences (and similarities) will manifest in terms of moral and social agency. We will require human points of reference in future HRI studies to fully understand how the emerging moral and social agency of robots relate to those qualities in humans.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

This paper was written by RJ with input, feedback, and review by TW.

FUNDING

This work was funded in part by Air Force Young Investigator Award 19RT0497.

ACKNOWLEDGMENTS

We would like to thank Dylan Doyle-Burke for his collaboration during earlier stages of this work.

REFERENCES

Alač, M. (2016). Social Robots: Things or Agents?. *AI Soc.* 31, 519–535.

Atkinson, R. K., Mayer, R. E., and Merrill, M. M. (2005). Fostering Social agency in Multimedia Learning: Examining the Impact of an Animated Agent's Voice. *Contemp. Educ. Psychol.* 30, 117–139. doi:10.1016/j.cedpsych.2004.07.001

Banks, J. (2019). A Perceived Moral agency Scale: Development and Validation of a Metric for Humans and Social Machines. *Comput. Hum. Behav.* 90, 363–371. doi:10.1016/j.chb.2018.08.028

Baxter, P., Kennedy, J., Vollmer, A.-L., de Greeff, J., and Belpaeme, T. (2014). "Tracking Gaze over Time in Hri as a Proxy for Engagement and Attribution of Social agency," in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, 3-6 March 2014 (IEEE), 126–127.

Billett, S. (2008). Learning throughout Working Life: a Relational Interdependence between Personal and Social agency. *Br. J. Educ. Stud.* 56, 39–58. doi:10.1111/j.1467-8527.2007.00394.x

Bonnefon, J.-F., Shariff, A., and Rahwan, I. (2016). The Social Dilemma of Autonomous Vehicles. *Science* 352, 1573–1576. doi:10.1126/science.aaf2654

Briggs, G., and Scheutz, M. (2014). How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress. *Int. J. Soc. Robotics* 6, 343–355. doi:10.1007/s12369-014-0235-1

Brown, P., and Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Castelfranchi, C. (1998). Modelling Social Action for Ai Agents. *Artif. intelligence* 103, 157–182. doi:10.1016/s0004-3702(98)00056-3

Danaher, J. (2020). Welcoming Robots into the Moral circle: a Defence of Ethical Behaviourism. *Sci. Eng. Ethics* 26, 2023–2049. doi:10.1007/s11948-019-00119-x

Dennett, D. C. (1978). "Three Kinds of Intentional Psychology," in *Perspectives in the Philosophy of Language: A Concise Anthology*, 163–186.

Dobres, M.-A., and Hoffman, C. R. (1994). Social agency and the Dynamics of Prehistoric Technology. *J. Archaeol Method Theor.* 1, 211–258. doi:10.1007/bf02231876

Floridi, L., and Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and machines* 14, 349–379. doi:10.1023/b:mind.0000035461.63578.9d

Floridi, L. (2008). The Method of Levels of Abstraction. *Minds & Machines* 18, 303–329. doi:10.1007/s11023-008-9113-7

Gardner, A. (2016). *Agency Uncovered: Archaeological Perspectives on Social agency, Power, and Being Human*. New York: Routledge).

Garibay, J. C. (2018). Beyond Traditional Measures of Stem success: Long-Term Predictors of Social agency and Conducting Research for Social Change. *Res. High Educ.* 59, 349–381. doi:10.1007/s11162-017-9470-2

Garibay, J. C. (2015). STEM Students' Social agency and Views on Working for Social Change: Are STEM Disciplines Developing Socially and Civically Responsible Students?. *J. Res. Sci. Teach.* 52, 610–632. doi:10.1002/tea.21203

Ghazali, A. S., Ham, J., Barakova, E., and Markopoulos, P. (2018). The Influence of Social Cues in Persuasive Social Robots on Psychological Reactance and Compliance. *Comput. Hum. Behav.* 87, 58–65. doi:10.1016/j.chb.2018.05.016

Ghazali, A. S., Ham, J., Markopoulos, P., and Barakova, E. I. (2019). "Investigating the Effect of Social Cues on Social agency Judgement," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, 11-14 March 2019 (IEEE), 586–587. doi:10.1109/HRI.2019.8673266

Gino, F. (2015). Understanding Ordinary Unethical Behavior: Why People Who Value Morality Act Immorally. *Curr. Opin. Behav. Sci.* 3, 107–111. doi:10.1016/j.cobeha.2015.03.001

Gray, K., and Wegner, D. M. (2009). Moral Typecasting: Divergent Perceptions of Moral Agents and Moral Patients. *J. Personal. Soc. Psychol.* 96, 505–520. doi:10.1037/a0013748

Heerink, M., Kröse, B., Evers, V., and Wielinga, B. (2010). Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. *Int. J. Soc. Robotics* 2, 361–375. doi:10.1007/s12369-010-0068-5

Himmelreich, J. (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethic Theor. Moral Prac* 21, 669–684. doi:10.1007/s10677-018-9896-4

Jackson, R. B., Wen, R., and Williams, T. (2019). "Tact in Noncompliance: The Need for Pragmatically Apt Responses to Unethical Commands," in AAAI Conference on Artificial Intelligence, Ethics, and Society, January 2019 (Honolulu, HI: ACM), 499–505. doi:10.1145/3306618.3314241

Jackson, R. B., and Williams, T. (2019). "Language-capable Robots May Inadvertently Weaken Human Moral Norms," in Companion Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI), Daegu, Korea (South), 11-14 March 2019 (IEEE). doi:10.1109/HRI.2019.8673123

Jackson, R. B., Williams, T., and Smith, N. (2020). "Exploring the Role of Gender in Perceptions of Robotic Noncompliance," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, March 2020 (Cambridge, United Kingdom: ACM), 559–567. doi:10.1145/3319502.3374831

Johnson, D. G., and Miller, K. W. (2008). Un-making Artificial Moral Agents. *Ethics Inf. Technol.* 10, 123–133. doi:10.1007/s10676-008-9174-6

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., et al. (2011). "The New Ontological Category Hypothesis in Human-Robot Interaction," in 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Lausanne, March 2011 (IEEE), 159–160. doi:10.1145/1957656.1957710

Kennedy, J., Baxter, P., and Belpaeme, T. (2014). "Children Comply with a Robot's Indirect Requests," in 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Bielefeld, 3-6 March 2014 (IEEE).

Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are Physically Embodied Social Agents Better Than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People's Loneliness in Human-Robot Interaction. *Int. J. human-computer Stud.* 64, 962–973. doi:10.1016/j.ijhcs.2006.05.002

Lee, M. K., Kiesler, S., Forlizzi, J., and Rybski, P. (2012). "Ripple Effects of an Embedded Social Agent: a Field Study of a Social Robot in the Workplace," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 2012 (Austin, TX: ACM), 695–704. doi:10.1145/2207676.2207776

Levin, D. T., Adams, J. A., Saylor, M. M., and Biswas, G. (2013). "A Transition Model for Cognitions about agency," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Tokyo, Japan, 3-6 March 2013 (IEEE), 373–380. doi:10.1109/HRI.2013.6483612

Luria, M., Hoffman, G., Megidish, B., Zuckerman, O., and Park, S. (2016). "Designing Vyo, a Robotic Smart home Assistant: Bridging the gap between Device and Social Agent," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, 26-31 Aug. 2016 (IEEE), 1019–1025. doi:10.1109/ROMAN.2016.7745234

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., and Wykowska, A. (2019). Do we Adopt the Intentional Stance toward Humanoid Robots?. *Front. Psychol.* 10, 450. doi:10.3389/fpsyg.2019.00450

Meyer, J. W., and Jepperson, R. L. (2000). The "Actors" of Modern Society: The Cultural Construction of Social Agency. *Sociological Theor.* 18, 100–120. doi:10.1111/0735-2751.00090

Nagao, K., and Takeuchi, A. (1994). Social Interaction: Multimodal Conversation with Social Agents. *AAAI* 94, 22–28.

Nass, C., Steuer, J., and Tauber, E. R. (1994). "Computers Are Social Actors," in Proceedings of the SIGCHI conference on Human factors in computing systems, April 1994 (Boston, MA: ACM), 72–78. doi:10.1145/191666.191703

Nomura, T., Uratani, T., Kanda, T., Matsumoto, K., Kidokoro, H., Suehiro, Y., et al. (2015). "Why Do Children Abuse Robots?," in HRI Extended Abstracts, March 2015 (Portland, OR: ACM), 63–64. doi:10.1145/2701973.2701977

Perez-Osorio, J., and Wykowska, A. (2019). "Adopting the Intentional Stance towards Humanoid Robots," in *Wording Robotics* (Springer), 119–136. doi:10.1007/978-3-030-17974-8_10

Pollini, A. (2009). A Theoretical Perspective on Social agency. *AI Soc.* 24, 165–171. doi:10.1007/s00146-009-0189-2

Roubroeks, M., Ham, J., and Midden, C. (2011). When Artificial Social Agents Try to Persuade People: The Role of Social agency on the Occurrence of Psychological Reactance. *Int. J. Soc. Robotics* 3, 155–165. doi:10.1007/s12369-010-0088-1

Schellen, E., and Wykowska, A. (2019). Intentional Mindset toward Robots-Open Questions and Methodological Challenges. *Front. Robot. AI* 5, 139. doi:10.3389/frobt.2018.00139

Straub, I. (2016). 'It Looks like a Human!' the Interrelation of Social Presence, Interaction and agency Ascription: a Case Study about the Effects of an Android Robot on Social agency Ascription. *AI Soc.* 31, 553–571. doi:10.1007/s00146-015-0632-5

Thellman, S., Silvervarg, A., and Ziemke, T. (2017). Folk-psychological Interpretation of Human vs. Humanoid Robot Behavior: Exploring the Intentional Stance toward Robots. *Front. Psychol.* 8, 1962. doi:10.3389/fpsyg.2017.01962

Thellman, S., and Ziemke, T. (2019). "The Intentional Stance toward Robots: Conceptual and Methodological Considerations," in The 41st Annual Conference of the Cognitive Science Society, Montreal, Canada, July 24-26, 1097–1103.

Ullman, D., Leite, L., Phillips, J., Kim-Cohen, J., and Scassellati, B. (2014). "Smart Human, Smarter Robot: How Cheating Affects Perceptions of Social agency," in Proceedings of the Annual Meeting of the Cognitive Science Society, Quebec, Canada.

Verbeek, P.-P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. University of Chicago Press.

Westlund, J. M. K., Martinez, M., Archie, M., Das, M., and Breazeal, C. (2016). "Effects of Framing a Robot as a Social Agent or as a Machine on Children's Social Behavior," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, 26-31 Aug. 2016 (IEEE), 688–693. doi:10.1109/ROMAN.2016.7745193

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer (AW) declared a past co-authorship with one of the authors (TW) to the handling Editor.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/frobt.2021.707149

Check for updates

Task-Level Authoring for Remote Robot Teleoperation

Emmanuel Senft[1], Michael Hagenow[2], Kevin Welsh[1], Robert Radwin[3], Michael Zinn[2], Michael Gleicher[1] and Bilge Mutlu[1]*

[1]*Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI, United States,* [2]*Department of Mechanical Engineering, University of Wisconsin–Madison, Madison, WI, United States,* [3]*Department of Industrial System and Engineering, University of Wisconsin–Madison, Madison, WI, United States*

Remote teleoperation of robots can broaden the reach of domain specialists across a wide range of industries such as home maintenance, health care, light manufacturing, and construction. However, current direct control methods are impractical, and existing tools for programming robot remotely have focused on users with significant robotic experience. Extending robot remote programming to end users, i.e., users who are experts in a domain but novices in robotics, requires tools that balance the rich features necessary for complex teleoperation tasks with ease of use. The primary challenge to usability is that novice users are unable to specify complete and robust task plans to allow a robot to perform duties autonomously, particularly in highly variable environments. Our solution is to allow operators to specify shorter sequences of high-level commands, which we call task-level authoring, to create periods of variable robot autonomy. This approach allows inexperienced users to create robot behaviors in uncertain environments by interleaving exploration, specification of behaviors, and execution as separate steps. End users are able to break down the specification of tasks and adapt to the current needs of the interaction and environments, combining the reactivity of direct control to asynchronous operation. In this paper, we describe a prototype system contextualized in light manufacturing and its empirical validation in a user study where 18 participants with some programming experience were able to perform a variety of complex telemanipulation tasks with little training. Our results show that our approach allowed users to create flexible periods of autonomy and solve rich manipulation tasks. Furthermore, participants significantly preferred our system over comparative more direct interfaces, demonstrating the potential of our approach for enabling end users to effectively perform remote robot programming.

Keywords: human-robot interaction, end-user programing, teleoperation, robotics, remote robot control, user study

OPEN ACCESS

Edited by:
Stefanos Nikolaidis, University of Southern California, United States

Reviewed by:
Xuning Yang, Carnegie Mellon University, United States
Yi-Shiuan Tung, University of Colorado Boulder, United States

****Correspondence:***
Emmanuel Senft
esenft@wisc.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *09 May 2021*
Accepted: *03 September 2021*
Published: *27 September 2021*

Citation:
Senft E, Hagenow M, Welsh K, Radwin R, Zinn M, Gleicher M and Mutlu B (2021) Task-Level Authoring for Remote Robot Teleoperation. Front. Robot. AI 8:707149. doi: 10.3389/frobt.2021.707149

1 INTRODUCTION

Effective teleoperation of robots—broadly, a remote human controlling a robot at a distance (Niemeyer et al., 2016)—is critical in scenarios where automation is impractical or undesirable. When a person operates a remote robot, they must acquire sufficient awareness of the robot's environment through sensors and displays, be able to make decisions about what the robot should do, provide directions (control) to the robot, and evaluate the outcomes of these operations. These challenges have been addressed with a wide range of interfaces that span a continuum of levels of autonomy (Beer et al., 2014), ranging from direct control where the operator drives the moment-to-moment details of a robot's movements, to asynchronous control, where operators send complex programs to the robot to execute autonomously, e.g., space exploration where robots receive programs for a day's worth of activities (Maxwell et al., 2005). The choices of level of control provide different trade-offs to address the goals of a specific scenario. In particular, longer-horizon control offers better robustness to communication issues and provides long periods of idle time for the operator while the robot is executing the commands. However, it also limits the opportunities for the human to react to unexpected situations during the program execution and requires significant huamn expertise to design robust behaviors and advanced sensing skills for the robot. On the other hand, more direct control allows operators to react quickly and easily to uncertainty, but demands constant attention from the operator, often relies on dedicated hardware, and requires a fast and stable connection to ensure that the tight real-time loop between the operator and the robot is maintained.

Our goal is to provide effective telemanipulation for end-user applications, such as home care, light manufacturing, or construction. In such scenarios, high level robot autonomy of autonomy would be desirable, as this would reduce the operator's workload, however there remains situations where a fully autonomous behavior cannot be created. Users have domain knowledge, they can analyze the robot environment and determine appropriate actions for the robot, but they have no expertise in creating robot programs. Building a system supporting teleoperation for these novice users presents a number of challenges, the system needs to 1) be easy to use, 2) support active perception (Bajcsy, 1988), 3) support specification of robot behaviors adapted to the current state of the environment, and 4) allow for periods of autonomy. As we will detail in **Section 2**, current interfaces for teleoperation are often specifically tailored to highly trained operators or adopt a low level of autonomy. The former are not suited to novice users and the latter forces users to continuously provide inputs to the robot, reducing both the usability over extended periods of time and increasing the sensitivity to communication issues.

Our key idea is to use task-level authoring to enable the operator to control the robot by specifying semantically connected sequences of high-level (task-level) steps. This paradigm supports various lengths of program depending on available environment information, ranging from single actions to longer plans. For example, a robot might need to open a drawer with a specific label and empty it, however the robot does not have character recognition. The operator could use the robot to locate the appropriate drawer and then create a plan for the robot to open this specific drawer, remove all items in it, and then close it. Task-level authoring aims to offer more flexibility for the operator, allowing them both to specify long periods of autonomy when possible, but also have a more direct control when necessary to allow the operator to obtain the environmental awareness necessary to make longer plans.

We propose four principles to support effective telemanipulation by novices:

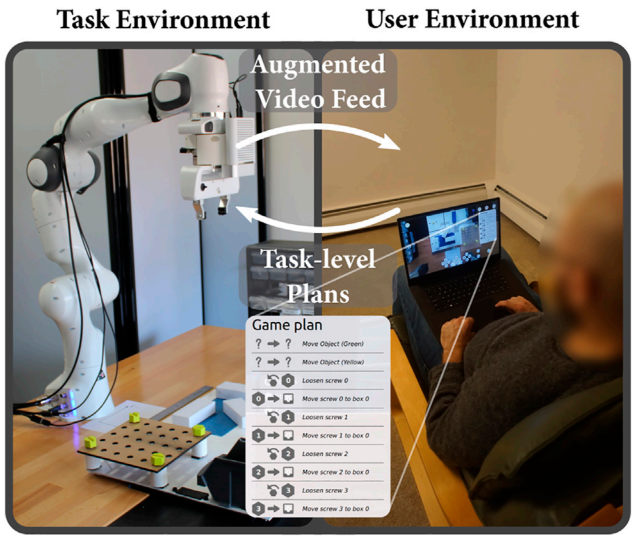


FIGURE 1 | Drawing Board: task-level authoring for robot teleoperation. By watching a robot-centric augmented video feed and annotating it, novice users can acquire awareness about the robot's environment and send task-level plans to control a robot remotely.

1) Interleaving observation and planning: the stepwise nature of manipulation tasks allows phases of observing the environment to gain awareness with phases of acting on that information. Execution occurs asynchronously, allowing it to be robust against communication problems and providing idle time to the user. Users can assess the state of the environment, devise a short plan for the robot, execute it, and the restart the process with the new state of the environment.
2) Controlling the robot at the action level: instead of controlling the robot motions, operators can select actions for the robot (e.g., pick-up, pull, or loosen). Such higher level of control allows participants to focus on the task that needs to be solved instead of the robot's kinemathics or workspace geometry.
3) Providing a unified augmented reality interface: task specification can be accomplished from a viewpoint chosen by the user to be convenient, as part of their awareness gathering process. This process allows us to use a screen overlay-based augmented reality interface that aggregates all the required information for decision making on a single view also used to specify actions. This single view makes the programming easier for the users as all the important information are available in a single place.
4) Specifying actions graphically: the augmented-reality interface allows for details of operations to be specified and verified graphically in context, simplifying the interface further. Additionally, such graphical specification allows easily to generalize a plan to a group of objects of the same type.

We have prototyped these ideas in a system called Drawing Board after an artist's portable drawing board (**Figure 1**) and evaluated it in a user study with 18 participants. Our central contribution is to show that a task-level authoring approach can be applied to teleoperation to create a system that affords both ease-of-use and asynchronous operation. In our study, remote operators (students with limited programming knowledge) were able to perform complex tasks, gaining the benefits of asynchronous operation (robustness to delays and opportunities for longer periods of idle time) with the ease-of-use and reactivity of more direct interfaces. Participants—some literally on the other side of the world—were able to teloperate the robot with little training and preferred our system compared to interfaces not embodying our principles. These findings show how the core choice of task-level authoring is supported by specific interface and implementation designs, yielding a system that meets our goals, allowing end users to remotely create short period of autonomy for robots.

2 RELATED WORK

Our work brings elements from the field of authoring and end-user programming to teleoperation.

2.1 Teleoperation

Fundamentally teleoperation refers to human control of robot actions, typically done remotely (i.e., the human and the robot are not collocated and the human can only perceive the robot's environment through artificial sensors and displays) (Niemeyer et al., 2016). With teleoperation, the question of the appropriate level of autonomy is important, especially in the presence of delay and partial situational awareness (Yanco et al., 2015; Niemeyer et al., 2016). Levels of autonomy form a continuum between direct control and long-term programs:

1) Direct control: low-level control where the operator is manually controlling all actions of the robot in real-time (e.g., remote surgery (Marescaux et al., 2001), military (Yamauchi, 2004)).
2) Semi-autonomy: the human operator intermittently controls robot actions where required and can parameterize higher-level actions that are executed autonomously by the robot (e.g., search-and-rescue, DARPA Robotics Challenge (Johnson et al., 2015)).
3) Teleprogramming: operators create programs defining actions and reactions to changes in the environment for the robot to execute over longer period of time (e.g., Mars rovers programmed every day for a full day of autonomy (Norris et al., 2005)).

Direct control has seen widespread use in the aerospace, nuclear, military, and medical domains (Niemeyer et al., 2016) as it allows operators to quickly react to new information. However, this type of teleoperation requires constant inputs from the operator and is highly sensitive to communications problems. Researchers have explored various methods to address this communication challenge. One direction of research involves optimizing the communication channel itself to reduce delay and allow the operator to have quick feedback on their actions (Preusche et al., 2006). Another method, shared control, seeks to make the process more robust to human error through means such as virtual fixture methods, which support the operator in their direct manipulation task (Rosenberg, 1993) or alternating phases of teleoperation and autonomous operation (Bohren et al., 2013). Finally, a third alternative uses a virtual model of the workspace to provide rapid feedback to the user from simulation while sending commands to the robot (Funda et al., 1992).

On the other end of the spectrum, traditional programming for autonomy and teleprogramming provides only limited feedback to the operator about the robot behavior. Operators need to have complete knowledge about the task including all required contingencies, to create dedicated programs for each task. These programs must be robust enough to run autonomously for hours without feedback. Furthermore, the robot needs to have the sensing capabilities to capture and analyze every relevant information in the environment. This highly autonomous control method is especially useful where there are large time delays between the robot and the operator which prevents the operator from intervening in real-time, such as when controlling a rover on Mars (Maxwell et al., 2005).

A semi-autonomous robot is a middle ground between these two extremes: it can execute short actions autonomously, but relies on the human operator to determine a plan of action and provide the correct parameters for these actions. The human (or team of humans) can use the robot to actively collect information about the environment, and provides near real-time inputs to the robot. The DARPA robotic challenge explores this space. In this case, the robot can run parameterized subroutines while multi-person teams of highly trained operators analyze data from the robot and control it at various abstraction-levels (from joint angle to locomotion goal), including situations with unstable communication channels (Johnson et al., 2015). These subroutines can be parameterized by selecting or moving virtual markers displaying the grasping pose (Kent et al., 2020), robot joint position (Nakaoka et al., 2014), or using affordance templates (Hart et al., 2014). In a retrospective analysis, Yanco et al. (2015) highlight the training required for operating the robots during these trials, and reports that researchers should explore new interaction methods that could be used by first responders without extensive training.

One approach to simplify both awareness acquisition and control (two keys aspects in teleoperation) is to use monitor-based augmented reality—overlaying digital markers on views from the real world (Azuma, 1997). For example, Schmaus et al. (2019) present a system where an astronaut in the space station controlled a robot on earth using this technology. Their point-and-click interface presents the video feed from the head camera of a humanoid robot with outlines of the detected objects and menus around the video. When clicking on one of these objects, the system filters the actions that can be done on this object to propose only a small subset of possible actions to the operator. Similarly, Chen et al. (2011) propose a multi-touch interface when actions are assigned to gestures on a video feed displayed on a touch screen. This type of point-and-click or gesture interface allows the remote operator to gain awareness about the environment and simply select high-level actions for the robot to perform. Simulations can also be overlaid with markers that users can manipulate to specify the desired position of a robot or its end-effector (Hashimoto et al., 2011; Hart et al., 2014; Nakaoka et al., 2014). However, despite these advances, little work has been done to explore and evaluate interfaces that allow naive operators to actively acquire awareness about the environment and create longer plans consisting of multiple actions.

2.2 Authoring

In the context of robotics, the term authoring refers to methods allowing end users to create defined robot behaviors (Datta et al., 2012; Guerin et al., 2015; Weintrop et al., 2018). The general process starts with a design period where an initial behavior is created, then the robot can be deployed in the real world and its behavior tested and refined in additional programming steps if needed. When the desired requirements are achieved, the authoring process is finished and the robot is ready to be deployed to interact autonomously. Authoring differs from classic programming in its focus on end users with limited background in computer sciences and seeks to address questions of how can these users design, or author, behaviors using modalities such as tangible interactions (Sefidgar et al., 2017; Huang and Cakmak, 2017), natural language (Walker et al., 2019), augmented- or mixed-reality (Cao et al., 2019a; Peng et al., 2018; Akan et al., 2011; Gao and Huang, 2019), visual programming environments (Glas et al., 2016; Paxton et al., 2017), or a mixture of modalities (Huang and Cakmak, 2017; Porfirio et al., 2019). Steinmetz et al. (2018) describe task-level programming as parameterizing and sequencing predefined skills composed of primitives to solve a task at hand. Their approach combines this task-level programming and programming by demonstration (Billard et al., 2008) to create manipulation behaviors.

While promising, classic authoring methods suffer from two limitations when applied to remote robot control. First, the authoring process is often considered as a single design step creating a fully autonomous behavior (Perzylo et al., 2016; Cao et al., 2019b). This monolithic approach differs from teleoperation which assumes that human capabilities (sensory or cognitive) are available at runtime to help the robot successfully complete a task. Second, many authoring methods such as PATI (Gao and Huang, 2019) or COSTAR (Paxton et al., 2017) use modalities only available in situations where the human operator and the robot are collocated (e.g., kinesthetic teaching, tangible interfaces, or *in-situ* mixed reality). For example, teach pendants—which are interfaces provided by manufacturers of industrial robots—are designed to be used next to the robot and often require the operator to manually move the robot. Consequently, while available to end users, such methods are not possible to use remotely. Our work is in the line of Akan et al. (2011), who used augmented reality to specify plans for a robotic arm. However, our premise is that to enable novice users to teleoperate robots, active perception (i.e., environment exploration) and behavior specification should be interleaved and coupled through a single simple interface, and that manipulation of graphic handles is a powerful way to specify parameters for actions.

3 DESIGN

To allow non-expert users to control robots remotely, we propose a system rooted in task-level authoring which allows users to navigate the live environment and specify appropriate robot behaviors. The following sections and **Figure 2** detail the key concepts of the system topology.

3.1 Interleaving Observation and Planning

Specifying full execution plans for a robot would allow to reduce the operator workload during plan execution, but requires significant expertise in robotics and highly capable robots. To allow end users to create adaptable periods of autonomy for the robot, we propose to use a task-level authoring approach. This approach simplifies the programming process by allowing the programmer to break tasks into sequences of high level actions based on what they observe at the moment. Users can chain together actions to create flexible periods of autonomy, adapted to their knowledge of the situation. For example, a set of actions may consist of grabbing a set of bolts in an area and moving them into a set grid pattern to fasten a structure. If the user is unsure what action is required next or if something unexpected occurs, the task-level authoring approach allows the user to explore the environment and create new programs based on the outcome of previous actions and new information.

Controlling robots at the task level creates a number of opportunities for end-user teleoperation; it allows the human to remain in the decision loop to provide necessary expertise, while maintaining an asynchronous workflow. Such design allows end-users to alternate between observing the environment, specifying robot actions, and executing sequences of commands. Operators can specify short actions to explore the environment by moving the robot camera, acquiring awareness, and selecting an appropriate view point to author task plans. Then, once they have gathered enough information about the environment to know their next actions, they can schedule a longer plan consisting on multiple actions to solve the current part of the task. This process can be repeated as much as needed which allows for plans to be tailored to the current state of the environment. The asynchronous execution also provides robustness to communication instability. The inclusion of the operator in the control loop takes away the complexity of teleprogramming by having the human making complex perceptions and decisions. Thus, it keeps the benefits of direct control without the requirement of a tight and stable control loop and maintain the benefits of asynchronous control without requiring to create complex programs and plan ahead for unknown future.

3.2 Controlling the Robot at the Action Level

As mentioned in **Section 2.1**, teloperation levels of control covers a spectrum from direct control to teleprogramming. Direct control can afford ease of use when the user is provided with intuive input device (Rakita et al., 2018), however it requires minute control from the operator and is very sensitive to delay.

As shown in Schmaus et al. (2019), controlling a robot at the action level provides a number of advantages for teleoperation. First, as actions are executed using a local control loop, it allows to be robust to delays in communication. Second, it is intuitive for users, new operators can pick-up the system easily without requiring the user to possess any knowledge about robotics and control. Nevertheless, controlling solely at the action level

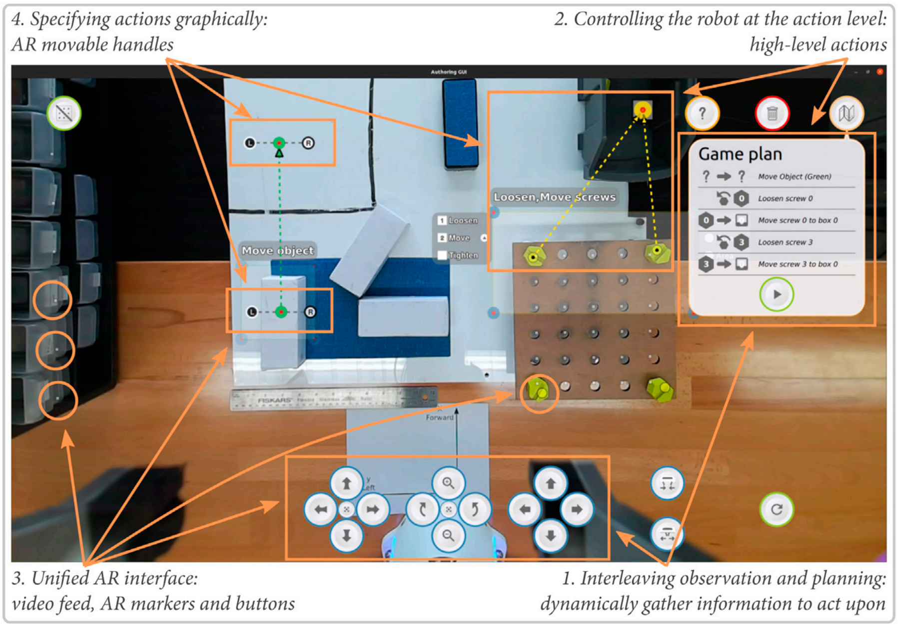


FIGURE 2 | Drawing Board's interface demonstrating our design principles.

suffers from some limitations. When using a single actions, even when users know what the robot should do over the next few actions, they have to specify an action, wait for it to be executed, specify the next action, and repeat, which can be suboptimal for the user. Additionally, similarly to any high-level control scheme, any action not in the robot vocabulary cannot be executed.

3.3 Providing a Unified, Augmented Reality Interface

Similar to some other robot authoring interfaces (Schmaus et al., 2019; Walker et al., 2019), our approach uses augmented reality (AR) to simplify perception and action specification. The interface is composed of a unified monitor-based AR interface showing a live camera view of the robot's environment augmented with digital markers representing detected objects (see **Figure 2**). The camera is mounted directly on the robot's end-effector for viewpoint flexibility and registration. The interface is overlaid with a canvas where the operator can design robot behaviors. This paradigm is consistent with research which shows that the most intuitive way to communicate information to an untrained operator is through vision (Yanco et al., 2004). More complex information such as the detected object pose and the environment point cloud are used to parameterize robot behavior behind the scenes, but hidden from the user's display.

3.4 Specifying Actions Graphically

One challenge in designing an interface for novice end users is to simplify the specification of complex manipulations. In programming, classic ways to set parameters are through sliders and numbers. Numerical parameter-setting allows greater precision, but can be unintuitive for users. Instead, our interface design leverages graphical representations whenever possible and minimizes required user input.

Our interface uses visual and interactive representations, mapped onto the augmented video feed, that enable users to parameterize predefined actions by manipulating these graphical representations. For example, to move a known object to a known positions, the interface creates anchors that can be moved by the user. Then, the interface will display an arrow from the starting point in the video to the goal point, visually representing the action in context. The interface uses 2D affordances throughout, as this is consistent with the 2D representation in video. 6D locations are inferred from the 2D interface based on environment information. Additionally, through graphical localization, a series of actions on a specific object can be generalized to nearby objects of the same category.

Our interface design only exposes high-level actions to the user (e.g., move, tighten, pull). The local robot controller decomposes these high-level actions into series of lower-level actions and translates them into primitives to reach the desired robot behavior. The user only has to specify the minimum fields required to execute the task and graphical specification allows to specify multiple parameters at the same time and in an intuitive matter.

4 IMPLEMENTATION

4.1 System

Following the considerations detailed in the previous section, we implemented Drawing Board, a prototype focused on enabling

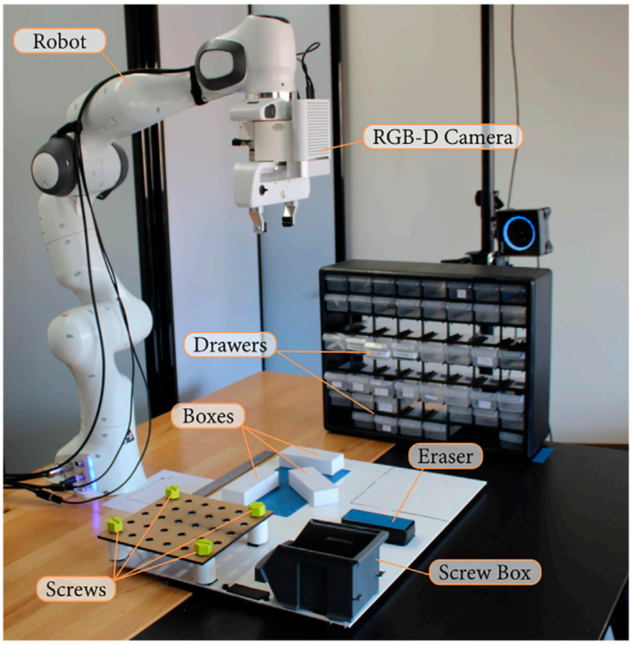


FIGURE 3 | Workspace used for our implementation.

users with little programming experience to operate a robot remotely. The interface was designed to be served from a traditional laptop or desktop display and focuses on controlling a single robotic manipulator. Our implementation integrates a collaborative robot (Franka Emika Panda) outfitted with an ATI Axia80-M20 6-axis force torque sensor and a Microsoft Azure Kinect providing both a 2D image and 3D point cloud. The camera is placed at the end-effector to allow for the greatest flexibility in camera position. The components of the system communicate using ROS (Quigley et al., 2009) with logic nodes implemented in Python, the graphical user interface in QML, and the low-level control in C++.[1]

For facilitating precise interaction with the environment, we implemented a hybrid controller to have more control over the forces applied by the robot when doing precise manipulation such as pulling a drawer. The hybrid control law follows an admittance architecture where interaction forces are measured from the force-torque sensor and resulting velocities are commanded in joint space via pseudo-inverse based inverse kinematics.

We also leverage the Microsoft Azure Kinect depth sensor to observe the environment. Objects are first localized in the scene by feeding the color image to Detectron2 (Wu et al., 2019), which provides a high fidelity binary pixel mask for each detected object. Once the object is localized, a GPU-accelerated Hough transform is used to register the known triangle mesh with each instance. This pipeline allows us to achieve 6D object pose estimation, which can then be used to provide the user with semantically correct actions as well as inform the robot motion plan. Our system also uses a number of predefined points of interest that represent the position of known static objects in the workspace. These known points are used as reference for the robot and to filter the position of objects detected by the live pose estimation pipeline.

4.2 Workspace

We applied our system to the workspace shown in **Figure 3**. This workspace guided our implementation, but the system can be adapted to other tasks or interactive objects. This workspace is composed of a number of drawers on the left of the robot with known positions. The middle of the workspace contains three white boxes above a blue area and a blue eraser, which are not detected by the robot object recognition system. The right part of the workspace contains a grid with holes and a screw box with known positions, and the grid can contain screws which are detected by the vision system.

The current prototype includes the following actions:

- Pulling and pushing the drawers;
- Picking, placing, and moving detected objects (e.g., screws) and undetected objects (e.g., boxes);
- Tightening and loosening the screws;
- Wiping an area.

By having general actions such as pull or move our system can adapt easily to other objects or different locations and orientations for these objects.

4.3 Interface

The default interface layout shows the video feed augmented with markers showing the detected or known points of interest (see **Figure 2**). The camera view is cropped to fill the full screen while showing clearly the robot's finger to allow users to know the gripper's status (open, closed, full).

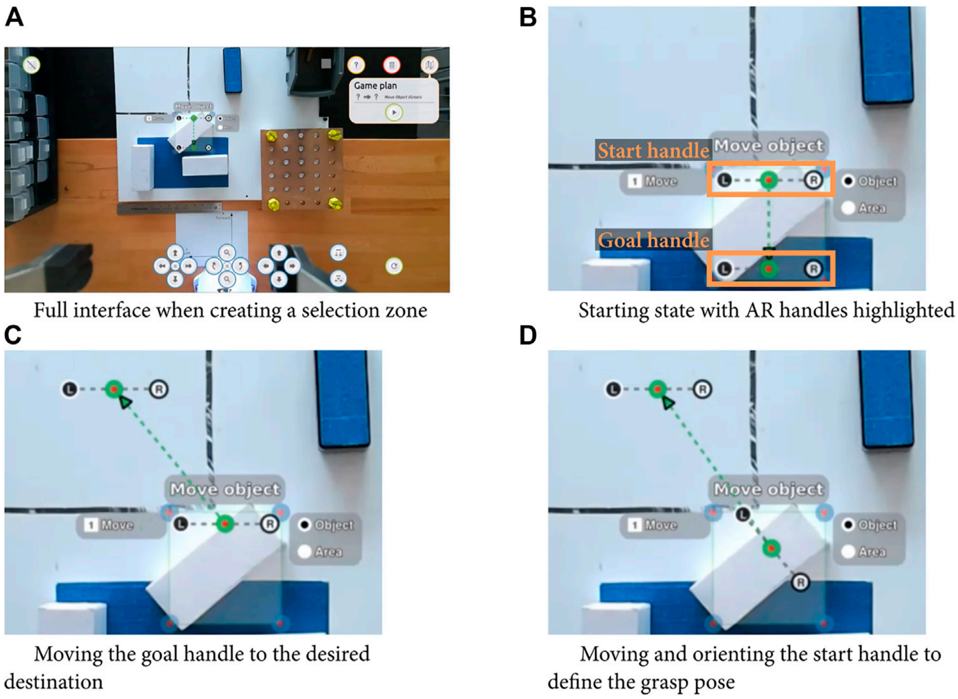
4.3.1 Direct Control

At the bottom of the screen there are a number of buttons for direct control: 12 buttons allow the user to move the camera by a discrete increment in each of the 6 potential directions (5 cm for the position buttons and $\pi/16$ radians for rotations), two buttons allow grasping and releasing, and a last button resets the robot to its homing position.

4.3.2 Authoring

To create task-level plans for the robots, users can annotate the augmented display to select actions applied to objects detected or parts of the environment. Users can click (or click-and-drag) on the screen to create selection areas to plan actions for the robot. Each selection area corresponds to one action or a set of actions on one type of object. Actions that can be parameterized (e.g., move actions) provide different types of handles that can be used to fully characterize the action. Users can create multiple selection areas to schedule different types of actions, and the resulting plan is shown in the Game Plan at the right of screen (see **Figure 2**). Users can use this game plan to confirm that the interface

[1]Open-source code for our system implementation is available at https://github.com/emmanuel-senft/authoring-ros/tree/study.



Full interface when creating a selection zone

Starting state with AR handles highlighted

Moving the goal handle to the desired destination

Moving and orienting the start handle to define the grasp pose

FIGURE 4 | Example of parameterization of a moving unknown object action, with full interface for the initial state and zoom in.

interpreted the intentions correctly before sending the plan to the robot. During the execution, the user can monitor the robot progress in the task by watching the video feed and checking the progress in the plan. Video examples can be found at https://osf.io/nd82j/.

4.3.3 Interaction with undetected objects

To pick and place an object not detected by the system, users can manipulate a start pose and a goal pose handles to specify the motion (see **Figure 4**). These handles are composed of three connected points: the interaction point (grasping or releasing) as well as points representing the robot's fingers, and users can move the handle on the screen to change the interaction location, and rotate it to specify the end-effector orientation. This pixel value is then mapped into a 3D point in the camera frame using the Kinect's depth camera and converted in a point in space for robot. The orientation from the interface specifies the rotation on the vertical axis and consequently completely characterize a vertical tabletop grasp.

4.3.4 Generalization to Groups of Objects

When creating a selection area, the interface will select a default object to interact with based on the ones present in the area, but the type of object can be changed by clicking on radio buttons displaying the objects present in the area. Each object has a number of actions that can be executed on it (e.g., a screw can be tightened, loosened, or moved), and the user can select which action to apply and in what order by using numbered checkboxes. These actions will then be applied on each object of the selected type in the area (e.g., loosen and move all the screws in the area).

4.4 Backend

The interface exposed the following high-level actions: move (both known and unknown objects), loosen, tighten, wipe, pull, and push. These high-level actions selected by users are grounded in the real world by finding the 6D pose of the interactions points in the user plan (either using the depth map from the Kinect or the location of points in a list of known objects). Each action is then hierarchically decomposed in a set of lower level actions (e.g., pick-up, view, place) and primitives (e.g., move to position, move to contact, grasps). For example, a move known object action is decomposed first into a pick and a place actions, which are then decomposed into a multiple of primitives (move above grasping point, move to grasping point, grasp, move above grasping point, move above release point, move to release point, release, and finally move above releasing point). During execution, the robot will perform each of the parameterized primitive to complete the plan.

This method can be extended to new applications in three ways: 1) adding new objects to the image recognition and interface affordances, 2) by composing existing primitives to create new higher-level actions, and 3) if needed, by creating new primitives. The first two improvements could be made using graphical interface without having to code (e.g., using approaches similar to Steinmetz et al. (2018)), however the third one would require actual code modification. This is similar to the current state-of-the-art cobots teach pendants: they expose a number of primitives that users can use to create behaviors, but any requirement not covered by the primitives (such as additional sensor-based interaction) would need code development to add the capability. Nevertheless, we could envisage a mixed system where creating new primitives and actions could be done locally, using learning from demonstration, and then exposed at a higher level to remote users using our interface.

5 EVALUATION

We conducted an evaluation to assess the impact of the design principles of our system. As mentioned in **Section 3**, our task-level authoring system is based around four principles: 1) interleaving observation and planning, 2) controlling the robot at the action level, 3) providing a unified, augmented reality interface, 4) graphical specification of actions. For the sake of the evaluation, the unified AR interface principle was not evaluated as too many different alternatives exist, however, we explored the three other axes. We conducted a 3×1 within-participants study to explore three types of interfaces embodying or not our design principles: our task-level authoring interface (TLA), a point-and-click interface (PC) inspired from Schmaus et al. (2019), and finally a Cartesian control interface (CC) as can be found traditionally on a cobot's teach pendant (e.g., PolyScope for the universal Robots[2]) or recent work in teleoperation (Marturi et al., 2016).

The CC condition does not use any of our design principles and serves as an alternative to kinesthetic teaching (Akgun et al., 2012) which cannot be applied due to the remote aspect and to direct control which would have required 6D input control on the user side. The PC condition only embodies the second design principle (controlling the robot at the action level). It corresponds to a simpler version of our interface, where the robot has similar manipulation capabilities (e.g., pick-up objects, loosen or tighten screws, pull drawers) but where actions can only be specified one at a time and where parameters have to be set numerically (e.g., using use sliders to specify parameters such as angles). The last condition TLA is the interface described in **Sections 3** and **4** and embodies all four of our principles.

The evaluation took place over Zoom,[3] a video conference platform, and we use the built-in remote screen control as a way to allow participants to control the robot from remote locations. We did not assess the latency inherent of such system, but estimated it around one second.

5.1 Hypotheses

Our evaluation uses the metrics *S* for the task score, a performance measure; *A* for robot autonomy, measured by both total and individual periods of autonomy; *U* for usability, measured by the SUS scale (Brooke, 1996); *P* for user preference for the control method; and *W* for workload, measured by the NASA Task-Load Index (NASA TLX) (Hart and Staveland, 1988). Below, we describe our hypotheses and provide specific predictions for each measure. Subscripts denote study conditions (TLA, PC, and CC).

[2]https://www.universal-robots.com

[3]https://zoom.us/

Our evaluation tested three hypotheses along the three evaluated design axes:

H1 Task score, autonomy, usability, and user preference will be higher, and workload will be lower with high-level control (PC, TLA) than low-level control (CC).
- Prediction P_{1a}: $S_{PC} > S_{CC}$, $A_{PC} > A_{CC}$, $U_{PC} > U_{CC}$, $P_{PC} > P_{CC}$, and $W_{PC} < W_{CC}$.
- Prediction P_{1b}: $S_{TLA} > S_{CC}$, $A_{TLA} > A_{CC}$, $U_{TLA} > U_{CC}$, $P_{TLA} > P_{CC}$, and $W_{TLA} < W_{CC}$.

H_2 Autonomy and user preference will be higher, and the workload will be lower when users are able to interleave observation and planning (TLA) than when they are not able to (PC).
- Prediction P_{2a}: $A_{TLA} > A_{PC}$.
- Prediction P_{2b}: $W_{TLA} < W_{PC}$.
- Prediction P_{2c}: $P_{TLA} > P_{PC}$.

H_3: Task score, usability, and user preference will be higher when users are able to parameterize actions graphically (TLA) than when they are not able to (PC).
- Prediction P_{3a}: $S_{TLA} > S_{PC}$.
- Prediction P_{3b}: $U_{TLA} > U_{PC}$.
- Prediction P_{3c}: $P_{TLA} > P_{PC}$.

*H*1 is based on the expectation that high-level action specification present in TLA and PC method automates away a large number of low-level actions that the user must specify in CC, which will save the user time and reduce the number of operations they must perform, thus their workload. *H*2 is grounded in the expectation that the task planning offered by our system will be used by participants to create longer periods of autonomy, which should reduce the workload, and make the participants prefer the method. Finally, *H*3 supposes that the graphical specification of actions will allow participants to specify action quicker (increasing their performance in the task), more easily (increasing the usability), and that participants will prefer this modality.

5.2 Method

5.2.1 Participants

We recruited 18 students enrolled in the Mechanical Engineering and Industrial and Systems Engineering departments at the university (3F/15M, age: $M = 19.6$, $SD = 1.54$). We selected our participants from this population as they represent people with some exposure but little expertise in robotics (familiarity with robots $M = 2.9$, $SD = 1.2$ on a five-point scale—none, a little, some, moderate, a lot—and familiarity with programming $M = 3$, $SD = 0.6$). The procedure was approved by the university's Institutional Review Board and participants were compensated at the rate of \$15/hour. The study was designed to last 80 min and included around 45 min of robot operation. Since it was completed remotely, participants stayed in their daily environment as shown in **Figure 5** top right, where a participant controlled the robot from his dorm bed.

5.2.2 Conditions

In all conditions, the layout of the interface was the same. It showed the camera feed, overlaid with arrows for direct control,

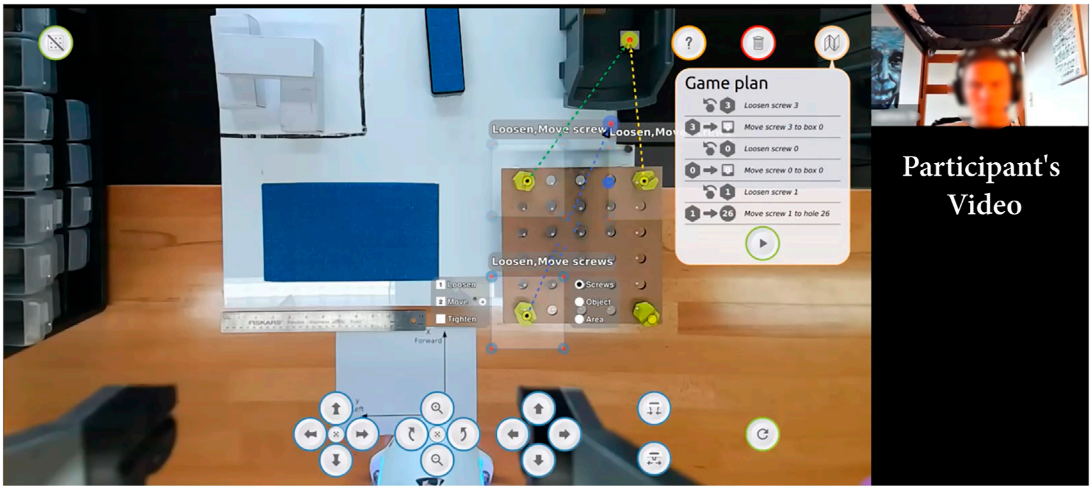


FIGURE 5 | Example of a participant using Drawing Board to control the robot from his dorm bed.

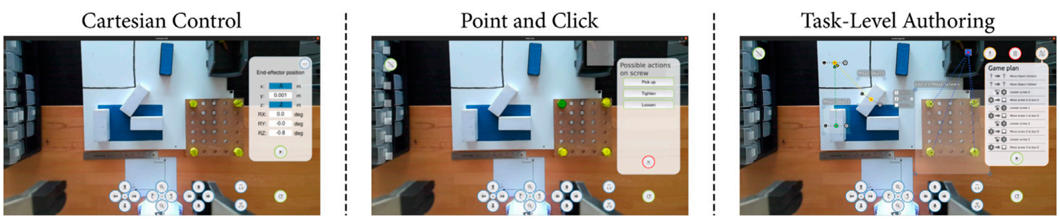


FIGURE 6 | Interfaces used in the study. Left shows the Cartesian Control interface: the user specifies numerically the end-effector position. Middle shows the Point-and-Click interface: the user selects direct actions on objects. Right shows the Task-Level Authoring interface allowing users to remotely create task plans for the robot.

buttons to grasp and release and the reset button. The difference was in the type of command sent to the robot as well as the modalities provided to the user. This study compared three conditions:

CC Cartesian control: the user uses six text boxes showing the current Cartesian position of the end-effector (x, y, z, rx, ry, rz) (see **Figure 6**-left). These text boxes can be edited with the desired command and sent to the robot either after modifying a single dimension or multiple ones.

PC Point-and-Click: the user is shown objects known by the system as markers overlayed on the video feed in an AR fashion (see **Figure 6**-center). The user can click on these markers or other parts of the view and is shown the different actions available on this object. Right clicking on an action allows to specify parameters, left clicking has the robot directly execute the action.

TLA Task-Level Authoring: interface presented in **Sections 3** and **4**, the user can annotate the video image to create actions associated to objects in the selection area and create task plans (see **Figure 6**-right).

To illustrate the difference between these conditions, we consider a move action on an unknown object (i.e., an object the operator can see, but the robot does not identify). With the CC condition, participants had to enter the 6D pose of a grasping point. Often, this process would be iterative, the operation would first have the robot approach the object by specifying a higher point, then correct the position and angle, then move to the grasping point. Then the operator had to press the grasp button, move to a dropping point (by specifying the 6D pose or using the camera control buttons), then press release. With the PC conditions, participants could click on the grasping location on the screen, parameterize the action with a grasp angle, execute the pick-up action, reset the robot, click on the destination location on screen and select the place action. With TLA, participant could click on the screen to create a section area, keep the move unknown object action (the default one if no identified object was in the selection area), move the start and goal handles (as shown in **Figure 4**), and finally press execute.

5.2.3 Tasks

As shown in **Figure 6**, the workspace has a number of drawers on the left, three white boxes on the bottom, an eraser at the top and four screws on the right.

Participants had to complete a training task followed by four additional tasks:

Task 0: Training - move the angled white box to the top left area.
Task 1: Pick - and - place - move the additional two boxes (at different orientations) to the top left area.
Task 2: Repeated actions - loosen the four screws from the grid and move them to the top - right gray box.
Task 3: Exploration - locate a specific drawer on the left, pull it, inspect its content and push it back.
Task 4: Continuous action - wipe the blue area with the eraser.

These tasks were selected to represent different types of actions that a remote operator may need to complete. The first three pick-and-place actions free the area that need to be wiped, demonstrating workspace manipulation actions. The loosening and moving of the four screws represent repeated actions. The drawer inspection task combines two awareness acquisition actions: locating the drawer and inspecting its content, as well as two workspace manipulation actions: opening and closing the drawer. Finally, the wiping action represents a continuous action over an area, similar to cleaning a table or sanding a piece.

Of note, the three pick-and-place actions (task 0 and task 1) requires the human to specify manually the grasping and placing point as the robot does not detect the boxes by itself. And the exploration (task 3) requires the operator to gather information outside of the default field of view by moving the camera on the robot, read the labels on the drawers, locate the relevant drawer, open the drawer, look into the drawer, count the numbers of items, and finally close the drawer. To be able to complete this task autonomously, a robot would need to have optical character recognition capabilities and be able to detect and count any type of object present in the drawers. Furthermore, as shown in **Algorithm 1**, if an operator wanted to design in a single step a program solving this task, the resulting program would require logic functions such as loops, conditional on sensors, functions within conditions, and loop breaking conditions. All these functionalities could be supported by more complex visual programming languages (such as Blockly[4]) which requires more knowledge in programming. Such a program would also require more capabilities for the robot, more complex representation of the world (e.g., having a list of all drawers with positions to read the label from, and positions to inspect the content), and more complex programming languages. Instead, using a human-in-the-loop approach (through direct control or task-level authoring) allows to achieve the same outcome, but with much simpler robot capabilities and interfaces.

```
for Each drawers, d do
    goToReadingPosition(d);
    l = readLabel(d);
    if l == "Bolt M5" then
        pull(d);
        goToInspectPosition(d);
        countObjects(d,"Bolt M5");
        push(d);
        break;
    else
        continue;
    end
end
```

Algorithm 1 | Example of algorithm to solve the exploration task autonomously.

5.2.4 Procedure

Participants joined a zoom call from their home or other daily environment. The study started with informed consent and a demographic questionnaire. Then participants were asked to watch a video introducing the robot and the workspace, followed by a second video introducing the tasks participants would need to complete [5]. For each condition, participants first watched a 2-min video presenting the main modalities of the interface and demonstrating how to make a pick-and-place action. Then, participants had 15 min to complete as many of the tasks as possible. During the training, they could ask any questions to the experimenter, however in the four later tasks the experimenter was only able to answer the most simple questions (e.g., "the screws are tightened down, right?" but not "how to move this object?").

The interaction with the robot stopped when participants reached 15 min or when they completed all the tasks. After this interaction, participants filled out NASA Task-Load Index (NASA TLX) (Hart and Staveland, 1988) and System Usability Scale (SUS) (Brooke, 1996) questionnaires before moving on to the next condition. The order of the conditions was counter-balanced and the study concluded with a semi-structured interview and a debriefing where participants could ask questions to the experimenter. Despite our best efforts, some participants created actions that could trigger the robot's safety locks (often due to excessive force being applied). In such situations, the timer was paused, the robot was restarted and the participant continued from where they stopped.

[4]https://developers.google.com/blockly

[5]All the videos are available at https://osf.io/nd82j/.

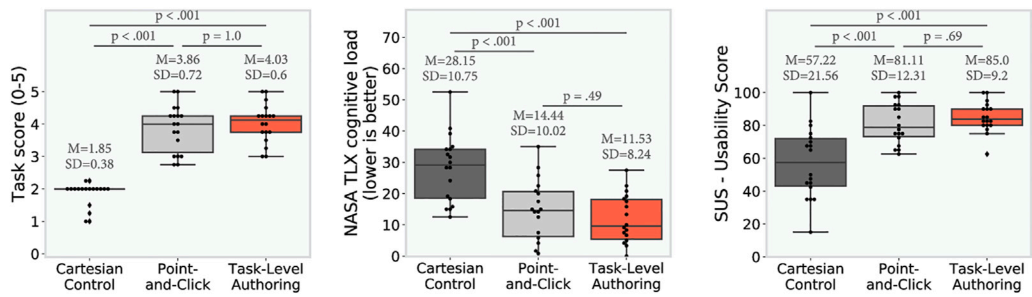


FIGURE 7 | Study results: *p*-values are computed using post-hoc paired *t*-test adjusted with Bonferroni correction (*n* = 18). Results show that both TLA and PC achieved a higher performance than CC (as shown by the task score), CC had a higher workload than both PC and TLA, and that both PC and TLA had a higher usability than CC. On the performance, workload, and usability no significant difference was observed between TLA and PC.

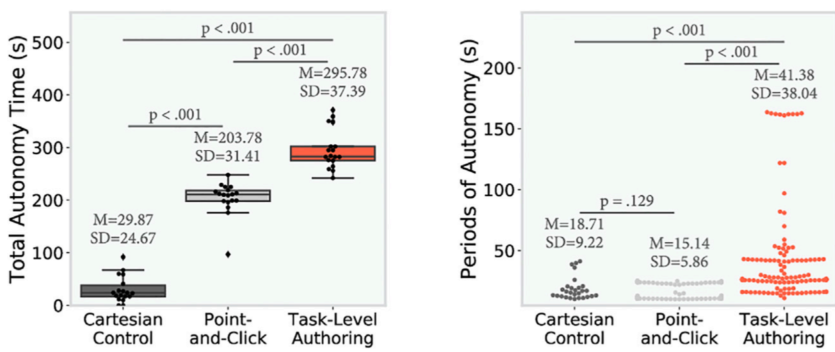


FIGURE 8 | Autonomy results. Left shows the total autonomy time for each conditions (only periods of autonomy of more than 10 s are counted), *p*-values are computed using post-hoc paired *t*-test adjusted with Bonferroni correction (*n* = 18). Right shows every single period of autonomy, *p*-values are computed using Games-Howell post-hoc to adjust for unequal sample size.

5.2.5 Measurements

We collected four types of quantitative data from the study. Task score, measured by how many tasks were fully or partially completed by the participants in the 15 min allocated per condition (with one point for the training and per task). Workload (i.e. how demanding it was to use the interface), measured by the NASA TLX. Usability (i.e. how intuitive the interface was), measured by SUS. Periods of autonomy, measured from period where the robot was moving continuously for more than 10 s (to be as inclusive as possible while not counting short periods that could barely be considered autonomous). We measured the periods of autonomy with a mixture of logs from the system and video coding of the interaction recordings using the Elan software (Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive, 2020).

In addition to the quantitative metrics, we collected qualitative impressions through the semi-structured interviews where we asked questions to the participants about their different experiences with the methods and which one they preferred.

6 RESULTS

Figure 7 and **Figure 8** present the quantitative results from the study. Results are first analyzed with a repeated measure ANOVA (corrected as needed using Greenhouse-Geisser), and then with post-hoc paired t-tests. A Bonferroni correction was directly applied to the *p*-values to protect against Type I error. For the periods of autonomy, as there was an unbalanced number of samples, we used ANOVA and Games-Howell post-hoc test.

6.1 Score

We observe significant impact of the condition on the score (sphericity was violated, Greenhouse-Geisser correction was used, $F\ (2, 34) = 115.53$, $p < 0.001$). Both the PC and the TLA interface achieved a score significantly higher than the CC (PC-CC: $t\ (17.0) = 13.0$, $p < 0.001$, TLA-CC: $t\ (17.0) = -23.0$, $p < 0.001$). However, we do not observe a significant difference of score between the PC and the TLA interfaces ($t\ (17.0) = -1.0$, $p = 1.0$).

Additionally, we did not observe an impact of the order, indicating that there no significant learning effect ($F\ (2, 34) = 0.698$, $p = 0.50$).

6.2 Periods of Autonomy

As shown in **Figure 8**, we observe a significant effect of the condition on the total autonomy time, $F(2, 34) = 297.45$, $p < 0.001$, and each condition was significantly different from the others, PC-CC: $t(17.0) = 18.0$, $p < 0.001$, TLA-CC: $t(17.0) = -25.0$, $p < 0.001$, TLA-PC: $t(17.0) = -7.0$, $p < 0.001$). With the CC offering the least amount of autonomy time, PC in the middle, and TLA offering the most of autonomy time. It can be observed that in addition to have a higher total autonomy time, the TLA condition also led to longer individual periods of autonomy than the other conditions (one way ANOVA: $F(2, 402) = 60.87$, $p < 0.001$, Games-Howell post-hoc test PC-CC: Mean Difference = −3.57, $p = 0.129$, TLA-CC: Mean Difference = −22.66, $p < 0.001$, TLA-PC: Mean Difference = −26.24, $p < 0.001$).

6.3 Workload

We observe significant effect of the condition on workload as measured by the NASA TLX, $F(2, 34) = 46.29$, $p < 0.001$. Both the PC and the TLA interface imposed a workload significantly lower than the CC (PC-CC: $t(17.0) = -9.0$, $p < 0.001$, TLA-CC: $t(17.0) = 8.0$, $p < 0.001$). However, we do not observe a significant difference of workload between the PC and the TLA interfaces ($t(17.0) = 1.0$, $p = 0.49$).

6.4 Usability

We observe a significant effect of the condition on usability as measured by the SUS, $F(2, 34) = 23.18$, $p < 0.001$. Both the PC and the TLA interface were rated as having a high usability (SUS score around 80) significantly outperforming the Cartesian interface (PC-CC: $t(17.0) = 4.0$, $p < 0.001$, TLA-CC: $t(17.0) = -6.0$, $p < 0.001$). However, we do not observe a significant difference of usability between the PC and the TLA interfaces ($t(17.0) = -1.0$, $p = 0.69$).

6.5 Preference

When asked which methods they preferred, 14 participants replied they preferred the TLA method, three preferred the PC method and one the CC method. Using one-sample binomial test, we measure a significant preference for our TLA method (95% Adjusted Wald Confidence Interval is (54.24%, 91.54%), preference TLA >33% with $p < 0.001$).

6.6 Observations and Feedback From Participants

In addition of our quantitative metrics, we made a number of anecdotal observation during the study and the following semi-structured interview. First, the two main justifications for participant's preference of the TLA interface were the ability to queue actions and the visualization (the two design principles not supported by the PC interface):

> "[TLA was] by far the best, because you could do so many tasks at once, and it was just really intuitive to figure out, okay, this is what it's gonna do"
>
> "I like that little line thing which would show up on positions, so you could determine like initial position and the final position [. . .] without having to remember numbers"
>
> "Being able to angle the jaws, and have visual reference for that, was really useful"

Some participants used the periods of autonomy of the TLA interface to perform secondary actions, e.g., drinking water or even as one participant did, sending a message to a friend. Combined to our quantitative results showing that the authoring interface frees longer periods of time to the operators, these observations provide anecdotal evidence that interfaces similar to TLA could help operators perform secondary action. However, as our study did not assess such an hypothesis, future work should confirm it.

Some participants were slightly confused by the different modalities used and monitored the game plan to understand how their inputs were parsed:

> "[TLA] also gave you that menu of like the order that you were going. I feel like that was really helpful"

Even though many participants qualified the PC interface of being very simple (almost too simple for some), participants still had to follow the progress of their series of action which can be complicated. For example, the screw task requires four repetitions of a loosen, a pick, a reset and a place action. Some participants in the study lost the track of which action was done, and for example forgot to loosen a screw, or did it twice. Some participants felt annoyed to have to re-specify each action every time:

> "[With TLA] you could I guess perform multiple tasks at once, you didn't have to click on it every single times"
>
> "[PC] it took more time, still because that you had to unscrew it, and then you had to pick it up, and then you had to move it and place it"

Being able to program the robots to sequence actions when having to repeat them over multiple objects allowed operators to keep track more easily of the progress in the task without having to keep in memory which actions were already executed.

Additionally, due to the lower granularity of control in CC and PC participants reported difficulties to know the distance between the robot's fingers and the table or objects or faced occlusion issues while operating the robot with the PC or CC interfaces. The task-level authoring offered by our system allowed participants to control the robot without facing these two obstacles.

7 DISCUSSION

7.1 Observations

Our results provide partial support for our hypotheses. H_1 is fully supported (both $P1_a$ and P_{1b} are supported). The higher level interfaces performed better than Cartesian Control on all metrics supporting H_1. H_2 is partially supported, TLA offered more autonomy than the other methods (supporting P_{2a}), and TLA was preferred to the other methods and participants referred to the ability of queuing actions as a reason (supporting P_{2c}), however TLA did not reduce the workload compared to PC (failing to support P_{2b}). Finally, H_3 is also partially supporting. Participants did not achieve a higher score with TLA than with PC and did not rate the usability as higher for TLA (failing to support both P_{3a} and P_{3b}), however, participants did prefer TLA over PC and participants referred to the graphical action parameterization as a reason (supporting P_{3c}).

Overall, these results show that our design principles partially achieved their goals: the high-level control allowed participants to think at the task level and progress quicker in the tasks. TLA was preferred overall due to the opportunity to create flexible periods of autonomy and the graphical parameterization of actions. This flexible programming horizon allowed participants to specify long periods of autonomy when possible, but also directly select actions when the next step is unclear. Traditionally, robots with more autonomy will require a lower workload at runtime, as the operator does not need to provided inputs when the robot is autonomous. However, such autonomous robots might inflict a higher workload at design time and require more skills for the operator and more capabilities for the robot. By interleaving exploration, design of short plans, and execution, TLA aims to maintain this low workload both at runtime and design time. Compared to specifying a behavior a priori, allowing the operator to specify commands at runtime allows to solve similar problems, but with simpler robot capabilities (as the operators can perform some sensory analysis) and simpler interface (as the operators does not have to create programs handling every possible situation).

We observed a potential ceiling effect on the usability (a score of 85 on the SUS is defined as excellent usability (Brooke, 2013)), and possibly a floor effect on the workload (14 and 11 on the NASA TLX are very low scores). These effects may have two distinct origins. Either our study was not sufficiently challenging for our operators, or our action sequencing principle did allow participants to obtain capabilities closer to programming (through the scheduling of action, automatic generalization of a set of actions to a group of objects etc.), which may have increased the complexity of the interface, but our graphical specification principle balanced this added complexity to maintained a low workload and high usability. Due to time constraints and study complexity, we could not explore individually the impact of each axis, which prevents us to identify the root cause of this effect. Future work should investigate more precisely the situations in which these methods could differ in usability and workload.

Nevertheless, from our study we can confirm that our authoring interface allowed participants to specify longer plans for the robot and streamlined the execution of repeated and composite actions. These two additional benefit might allow operators to perform secondary task, and potentially facilitate extended use (as anecdotally supported by the observation that some participants lost track of their progress in the repeated action and did the same action twice, or forgot whether they unscrewed a bolt already). This additional gain comes at no cost in term of workload and usability, which supports the conclusion that our design principles allowed the interface to be usable with limited training while incorporating additional programming capabilities. Future work should evaluate whether such increase in autonomy could allow operators to perform secondary tasks in practice and how such programming capabilities could be used by operators.

7.2 Limitations

Our approach suffers from a number of limitations that we plan to address in future work. A key limitation is that the high-level interface requires the specific primitives and actions to be pre-determined and pre-programmed. Extending the set of operations to support a broader range of tasks may create challenges in helping the user understand the range of options. Allowing users to specify actions that are not in the interfaces "vocabulary" is challenging, as this requires detailed specification that often must consider low-level control issues such as compliance. This issue is common in authoring—for example, teach pendants are also intrinsically limited in the robot's capabilities they expose and more complex uses often require coding. Additionally, our system relied on robust actions and we did not explore how to recover from failures when executing actions. We identify four ways such action failures could be handled. First, actions could be made more robust by integrating replanning strategies (e.g., planning a new grasp pose after a failed grasp). Second, high-level actions could take more parameters after a failure (e.g., specifying a full 6D grasp pose if the default one did not work). Third, the operator could provide additional runtime inputs to address small trajectory errors (Hagenow et al., 2021). And finally, the user could change the control mode for such infrequent event (e.g., temporarily using direct control instead of TLA).

The evaluation of our approach also has a number of limitations. For example, the study considered relatively simple tasks, used a mostly male population, and our population was not total novice, but had some experience with programming. Additionally, due to time constraints, we could not explore every single design axis individually. Future work should involve ablation studies, where the specific impact of design principles are evaluated, explore interactions in real environments, and use operators from the targeted population (family member controlling a robot in a home-assistant scenario or workers in a factory). Finally, future work should also explicitly explore the impact of latency when performing task-level authoring, especially compared to more direct control. We plan to address such limitations in future work.

7.3 Implications

Results from our evaluation lead a number of implications. Centrally, the use of task-level authoring seems to be an interesting trade-off, allowing for sufficient programming to gain the advantages of asynchronous control (i.e., programming longer periods of autonomy for the robot and leading to longer and better quality idle time, offloading some tasks following to the robot), yet having the programming be simple enough that it can be used during the interaction with little training. The approach affords an interface design that combines exploration, specification, and monitoring in a single view. The specific interface provides other general lessons. First, our work expands on the ideas of using higher-level controls to enable effective teleoperation interfaces. While prior systems have shown point-and-click interfaces (Schmaus et al., 2019), ours expands the concept to accomplish longer autonomous behavior. Second, by connecting these higher-level controls in a paradigm where exploration and manipulation are interleaved, we can create single-view interfaces that are usable in more complex scenarios. Third, our work extends prior see-through interfaces with camera control, allowing them to work in more environments. Fourth, our work shows the potential of asynchronous interfaces by improving the amount and duration of the offered periods of autonomy. By allowing the user to quickly specify longer plans, they gain opportunities for idle time, potentially freeing them to perform other tasks during execution. Finally, by demonstrating effective telemanipulation only using consumer interfaces shows that remote robot operation is possible for novice users—even at distances of many time zones.

8 CONCLUSION

In this paper, we explored the design of interfaces for remote control of a robotic arm by novice users. Our design considers the key goals of teleoperation interfaces: allowing remote novice operators to analyze the robot's environment and specify robot behavior appropriate to the situation. To address these challenges for scenarios with novice users and standard input devices we adopted a task-level authoring approach. The approach allowed for the design of an interface that interleaves exploration and planning, allowing us to utilize both direct control (more intuitive interface and benefiting from the human knowledge more frequently) and asynchronous control (robustness to communications issues and increased idle time for the operator). Our interface uses graphical overlays on a video feed of the environment to provide for simple exploration, specification of operations, and sequencing of commands into short programs. We evaluated a prototype system in an 18-participant study which showed that our interface allowed users with some familiarity with programming to 1) operate the robot remotely to gain awareness about the environment, 2) perform manipulation of the workspace, and 3) use the scheduling of actions to free long periods of idle times that might be used to perform secondary tasks. Furthermore, our interface was largely preferred compared to two other simpler interfaces.

Our work adds a new tool to the existing library of teleoperation approaches and demonstrates that task-level authoring is a powerful method to allow non-experts to remotely create short periods of autonomy for robots while allowing them to explore the robot's environment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Wisconsin-Madison Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ES, BM, and MG contributed to conception and design of the study. ES, KW, and MH designed and implemented the prototype. ES ran the study and performed the statistical analysis. ES wrote the first draft of the manuscript. MH, KW, MG, and BM wrote sections of the manuscript. BM, MG, RR, and MZ obtained funding for the project. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by a NASA University Leadership Initiative (ULI) grant awarded to the UW-Madison and The Boeing Company (Cooperative Agreement # 80NSSC19M0124).

REFERENCES

Akan, B., Ameri, A., Cürüklü, B., and Asplund, L. (2011). Intuitive Industrial Robot Programming through Incremental Multimodal Language and Augmented Reality. In 2011 IEEE International Conference on Robotics and Automation (IEEE), 3934–3939. doi:10.1109/icra.2011.5979887

Akgun, B., Cakmak, M., Yoo, J. W., and Thomaz, A. L. (2012). Trajectories and Keyframes for Kinesthetic Teaching: A Human-Robot Interaction Perspective. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. 391–398.

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments* 6, 355–385. doi:10.1162/pres.1997.6.4.355

Bajcsy, R. (1988). Active Perception. *Proc. IEEE* 76, 966–1005. doi:10.1109/5.5968

Beer, J. M., Fisk, A. D., and Rogers, W. A. (2014). Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *J. human-robot interaction* 3, 74. doi:10.5898/jhri.3.2.beer

Billard, A., Calinon, S., Dillmann, R., and Schaal, S. (2008). *Survey: Robot Programming by Demonstration*. Berlin: Springer Handbook of Robotics. doi:10.1007/978-3-540-30301-5_60

Bohren, J., Papazov, C., Burschka, D., Krieger, K., Parusel, S., Haddadin, S., et al. (2013).A Pilot Study in Vision-Based Augmented Telemanipulation for Remote Assembly over High-Latency Networks. In 2013 IEEE International Conference on Robotics and Automation. IEEE, 3631–3638. doi:10.1109/icra.2013.6631087

Brooke, J. (2013). Sus: a Retrospective. *J. usability Stud.* 8, 29–40.

Brooke, J. (1996). Sus: a "Quick and Dirty"usability. *Usability Eval. industry*, 189.

Cao, Y., Wang, T., Qian, X., Rao, P. S., Wadhawan, M., Huo, K., et al. (2019a). Ghostar: A Time-Space Editor for Embodied Authoring of Human-Robot Collaborative Task with Augmented Reality. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. 521–534. doi:10.1145/3332165.3347902

Cao, Y., Xu, Z., Li, F., Zhong, W., Huo, K., and Ramani, K. (2019b). V. Ra: An *In-Situ* Visual Authoring System for Robot-Iot Task Planning with Augmented Reality. In Proceedings of the 2019 on Designing Interactive Systems Conference. 1059–1070.

Chen, H., Kakiuchi, Y., Saito, M., Okada, K., and Inaba, M. (2011).View-based Multi-Touch Gesture Interface for Furniture Manipulation Robots. In Advanced Robotics and its Social Impacts. IEEE, 39–42. doi:10.1109/arso.2011.6301979

Datta, C., Jayawardena, C., Kuo, I. H., and MacDonald, B. A. (2012).Robostudio: A Visual Programming Environment for Rapid Authoring and Customization of Complex

Services on a Personal Service Robot. In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2352–2357. doi:10.1109/iros.2012.6386105

Funda, J., Lindsay, T. S., and Paul, R. P. (1992). Teleprogramming: Toward Delay-Invariant Remote Manipulation. *Presence: Teleoperators & Virtual Environments* 1, 29–44. doi:10.1162/pres.1992.1.1.29

Gao, Y., and Huang, C.-M. (2019). Pati: a Projection-Based Augmented Table-Top Interface for Robot Programming. In Proceedings of the 24th international conference on intelligent user interfaces. 345–355.

Glas, D. F., Kanda, T., and Ishiguro, H. (2016).Human-robot Interaction Design Using Interaction Composer Eight Years of Lessons Learned. In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 303–310. doi:10.1109/hri.2016.7451766

Guerin, K. R., Lea, C., Paxton, C., and Hager, G. D. (2015).A Framework for End-User Instruction of a Robot Assistant for Manufacturing. In 2015 IEEE international conference on robotics and automation (ICRA). IEEE, 6167–6174. doi:10.1109/icra.2015.7140065

Hagenow, M., Senft, E., Radwin, R., Gleicher, M., Mutlu, B., and Zinn, M. (2021). Corrective Shared Autonomy for Addressing Task Variability. *IEEE Robot. Autom. Lett.* 6, 3720–3727. doi:10.1109/lra.2021.3064500

Hart, S., Dinh, P., and Hambuchen, K. (2014). Affordance Templates for Shared Robot Control.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol. (Elsevier)* 52, 139–183. doi:10.1016/s0166-4115(08)62386-9

Hashimoto, S., Ishida, A., Inami, M., and Igarashi, T. (2011). Touchme: An Augmented Reality Based Remote Robot Manipulation. In The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011. vol 2.

Huang, J., and Cakmak, M. (2017).Code3: A System for End-To-End Programming of mobile Manipulator Robots for Novices and Experts. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 453–462.

Johnson, M., Shrewsbury, B., Bertrand, S., Wu, T., Duran, D., Floyd, M., et al. (2015). Team IHMC's Lessons Learned from the DARPA Robotics Challenge Trials. *J. Field Robotics* 32, 192–208. doi:10.1002/rob.21571

Kent, D., Saldanha, C., and Chernova, S. (2020). Leveraging Depth Data in Remote Robot Teleoperation Interfaces for General Object Manipulation. *Int. J. Robotics Res.* 39, 39–53. doi:10.1177/0278364919888565

Marescaux, J., Leroy, J., Gagner, M., Rubino, F., Mutter, D., Vix, M., et al. (2001). Transatlantic Robot-Assisted Telesurgery. *Nature* 413, 379–380. doi:10.1038/35096636

Marturi, N., Rastegarpanah, A., Takahashi, C., Adjigble, M., Stolkin, R., Zurek, S., et al. (2016). Towards Advanced Robotic Manipulation for Nuclear Decommissioning: A Pilot Study on Tele-Operation and Autonomy. In 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA). IEEE, 1–8. doi:10.1109/raha.2016.7931866

Maxwell, S., Cooper, B., Hartman, F., Leger, C., Wright, J., and Yen, J. (2005). The Best of Both Worlds: Integrating Textual and Visual Command Interfaces for mars Rover Operations. In 2005 IEEE International Conference on Systems, Man and Cybernetics (IEEE). 2, 1384–1388.

Nakaoka, S., Morisawa, M., Kaneko, K., Kajita, S., and Kanehiro, F. (2014).Development of an Indirect-type Teleoperation Interface for Biped Humanoid Robots. In 2014 IEEE/SICE International Symposium on System Integration. IEEE, 590–596.

Niemeyer, G., Preusche, C., Stramigioli, S., and Lee, D. (2016).Telerobotics. In Springer handbook of robotics. Springer, 1085–1108. doi:10.1007/978-3-319-32552-1_43

Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive (2020). Elan (version 5.9) [computer software].

Norris, J. S., Powell, M. W., Fox, J. M., Rabe, K. J., and Shu, I.-H. (2005).Science Operations Interfaces for mars Surface Exploration. In 2005 IEEE International Conference on Systems, Man and Cybernetics, 2. IEEE, 1365–1371.

Paxton, C., Hundt, A., Jonathan, F., Guerin, K., and Hager, G. D. (2017).Costar: Instructing Collaborative Robots with Behavior Trees and Vision. In 2017 IEEE international conference on robotics and automation (ICRA). IEEE, 564–571. doi:10.1109/icra.2017.7989070

Peng, H., Briggs, J., Wang, C.-Y., Guo, K., Kider, J., Mueller, S., et al. (2018). Roma: Interactive Fabrication with Augmented Reality and a Robotic 3d Printer. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 1–12.

Perzylo, A., Somani, N., Profanter, S., Kessler, I., Rickert, M., and Knoll, A. (2016).Intuitive Instruction of Industrial Robots: Semantic Process Descriptions for Small Lot Production. In 2016 ieee/rsj international conference on intelligent robots and systems (iros). IEEE, 2293–2300. doi:10.1109/iros.2016.7759358

Porfirio, D., Fisher, E., Sauppé, A., Albarghouthi, A., and Mutlu, B. (2019). Bodystorming Human-Robot Interactions. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. 479–491. doi:10.1145/3332165.3347957

Preusche, C., Reintsema, D., Landzettel, K., and Hirzinger, G. (2006).Robotics Component Verification on ISS ROKVISS-Preliminary Results for Telepresence. In 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 4595–4601. doi:10.1109/iros.2006.282165

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. (2009).Ros: an Open-Source Robot Operating System. In ICRA workshop on open source software, 3. Japan: Kobe, 5.

Rakita, D., Mutlu, B., and Gleicher, M. (2018). An Autonomous Dynamic Camera Method for Effective Remote Teleoperation. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 325–333. doi:10.1145/3171221.3171279

Rosenberg, L. B. (1993).Virtual Fixtures: Perceptual Tools for Telerobotic Manipulation. In Proceedings of IEEE virtual reality annual international symposium. IEEE, 76–82. doi:10.1109/vrais.1993.380795

Schmaus, P., Leidner, D., Krüger, T., Bayer, R., Pleintinger, B., Schiele, A., et al. (2019). Knowledge Driven Orbit-To-Ground Teleoperation of a Robot Coworker. *IEEE Robotics Automation Lett.* 5, 143–150.

Sefidgar, Y. S., Agarwal, P., and Cakmak, M. (2017).Situated Tangible Robot Programming. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction HRI. IEEE, 473–482. doi:10.1145/2909824.3020240

Steinmetz, F., Wollschläger, A., and Weitschat, R. (2018). RAZER-A HRI for Visual Task-Level Programming and Intuitive Skill Parameterization. *IEEE Robot. Autom. Lett.* 3, 1362–1369. doi:10.1109/lra.2018.2798300

Walker, M. E., Hedayati, H., and Szafir, D. (2019a).Robot Teleoperation with Augmented Reality Virtual Surrogates. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 202–210. doi:10.1109/hri.2019.8673306

Walker, N., Peng, Y.-T., and Cakmak, M. (2019b). *Robot World Cup*. Springer, 337–350. doi:10.1007/978-3-030-35699-6_26Neural Semantic Parsing with Anonymization for Command Understanding in General-Purpose Service Robots

Weintrop, D., Afzal, A., Salac, J., Francis, P., Li, B., Shepherd, D. C., et al. (2018). Evaluating Coblox: A Comparative Study of Robotics Programming Environments for Adult Novices. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 1–12.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. available at: https://github.com/facebookresearch/detectron2.

Yamauchi, B. M. (2004). Packbot: a Versatile Platform for Military Robotics. In Proc. SPIE 5422, Unmanned Ground Vehicle Technology VI, September 2, 2004, 228–237. doi:10.1117/12.538328

Yanco, H. A., Norton, A., Ober, W., Shane, D., Skinner, A., and Vice, J. (2015). Analysis of Human-Robot Interaction at the Darpa Robotics challenge Trials. *J. Field Robotics* 32, 420–444. doi:10.1002/rob.21568

Yanco, H., Drury, J., and Scholtz, J. (2004). Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition. *Human-comp. Interaction* 19, 117–149. doi:10.1207/s15327051hci1901&2_6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

SYSTEMATIC REVIEW
published: 19 November 2021
doi: 10.3389/frobt.2021.744526

10 Years of Human-NAO Interaction Research: A Scoping Review

Aida Amirova[1], Nazerke Rakhymbayeva[2], Elmira Yadollahi[3], Anara Sandygulova[2] and Wafa Johal[4,5]**

[1]Graduate School of Education, Nazarbayev University, Nur-Sultan, Kazakhstan, [2]Department of Robotics and Mechatronics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan, Kazakhstan, [3]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [4]University of Melbourne, Melbourne, VIC, Australia, [5]UNSW, Sydney, NSW, Australia

The evolving field of human-robot interaction (HRI) necessitates that we better understand how social robots operate and interact with humans. This scoping review provides an overview of about 300 research works focusing on the use of the NAO robot from 2010 to 2020. This study presents one of the most extensive and inclusive pieces of evidence on the deployment of the humanoid NAO robot and its global reach. Unlike most reviews, we provide both qualitative and quantitative results regarding how NAO is being used and what has been achieved so far. We analyzed a wide range of theoretical, empirical, and technical contributions that provide multidimensional insights, such as general trends in terms of application, the robot capabilities, its input and output modalities of communication, and the human-robot interaction experiments that featured NAO (e.g. number and roles of participants, design, and the length of interaction). Lastly, we derive from the review some research gaps in current state-of-the-art and provide suggestions for the design of the next generation of social robots.

Keywords: social robot, human-robot interaction, nao, survey, review, humanoid robot, qualitative, quantitative

OPEN ACCESS

Edited by:
Séverin Lemaignan, Pal Robotics S.L., Spain

Reviewed by:
Matthew Gombolay, Georgia Institute of Technology, United States
Bin Fang, Tsinghua University, China

****Correspondence:***
Anara Sandygulova
anara.sandygulova@nu.edu.kz
Wafa Johal
wafa.johal@unimelb.edu.au

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *20 July 2021*
Accepted: *21 October 2021*
Published: *19 November 2021*

Citation:
Amirova A, Rakhymbayeva N, Yadollahi E, Sandygulova A and Johal W (2021) 10 Years of Human-NAO Interaction Research: A Scoping Review. Front. Robot. AI 8:744526. doi: 10.3389/frobt.2021.744526

1 INTRODUCTION

For some decades, social robots have been used for research purposes in an attempt to assist humans and bring social benefits to their life. These social robots have been envisioned to interact with humans in various application domains such as education, healthcare, industry, entertainment, and public service. However, in order to claim that social robots reached their full potential as social assistive agents, they have to be able to create sustainable and intelligent interactions in the real world while acting in an acceptable and credible way. Therefore, the field of human-robot interaction has fueled research into the design, development and evaluation of social robots. There is a significant number of social robots in research, such as Kaspar for autism therapy (Wood et al., 2019), iCub for cognitive development (Natale et al., 2016), and Robovie for public spaces (Das et al., 2015), and the NAO robot. NAO has been among the most widely used social robots in human-robot interaction research due to its affordability and broad functionality. Developed by the French company, Aldebaran Robotics, in 2008 and acquired by the Japanese company, Softbank Robotics, in 2015, NAO is an autonomous and programmable humanoid robot that has been successfully applied to research and development applications for children, adults, and the elderly people. More than 13,000 NAO robots are used in more than 70 countries around the world. Consequently, a number of recent large-scale interdisciplinary projects, such as ALIZ-E[1], DREAM[2], CoWriter[3], SQUIRREL[4], L2Tor[5] have explored child-centered research with the mission to enable NAO to take a role of a tutor, a therapist, or a peer learner.

There have been several reviews about social robots used for specific application domains, such as robot-assisted education (Mubin et al., 2013; Belpaeme et al., 2018; Johal, 2020) and autism therapy (Saleh et al., 2020). There is evidence that NAO was among the heavily used social robots for these applications (Belpaeme et al., 2018; Saleh et al., 2020; Henschel et al., 2021). Among the most recent literature surveys, Robaczewski et al. (2020) reviewed the use of NAO as a socially assistive robot (SAR). The authors studied a total of 51 user-study publications and discussed their major findings around six themes: social engagement, affectivity, intervention, assisted teaching, mild cognitive impairment/dementia, and autism/intellectual disability. While providing a good overview of some of the social assistive robotics studies that were conducted with the NAO, this previous survey does not consider technical contributions, thus is limited in identifying research and development trends in its deployment across application domains. Therefore, it is still unclear how and why this social robot has been used in research over the last 10 years and how this standardized platform contributed more widely to the field of human-robot interaction.

For these reasons, a scoping review was a necessary step to systematically map the research done with the NAO robot in HRI and identify research trends and potential gaps of investigations that could lead to the development of a new standard platform for social robotics research. It seems a worthwhile effort to reflect on the dynamics of the socially acceptable robot - a humanoid NAO robot - that has a particular appeal for improving the social, behavioral, physical, and cognitive well-being of humans of various age groups. The present paper aims to provide a holistic understanding of the NAO robot for research by analyzing the unrestricted type of contributions, both theoretical and experimental. We also report on technical contributions that helped the field of HRI to grow over the years. While following a strict and reproducible protocol, our review probably does not cover the complete literature work in HRI research with the NAO robot. However, we consider that our screening protocol allowed to capture a good amount of the body of research using NAO and to present useful insights, findings, and trends in the use of the robot in the past decade. Unlike previous reviews, our research approach allows us to present general and specific findings that were gleaned from quantitative and qualitative analysis. We find our review vital in understanding how the social robots like NAO serve educational, professional, and social roles when interacting with humans and what are the crucial insights about its use and prospects. This research potentially benefits a wider community of stakeholders such as novice and expert HRI researchers, robotics labs or startups and those professionals working at the intersection of interdisciplinary fields like education and healthcare.

Our meta-analysis seeks to provide broad insights into the use of NAO in HRI by annotating a wide range of categories of applications (including but not limited to social assistive robotics), geographical distribution, type of contribution, application fields, experimental methodology, duration, and the number of sessions, human-robot ratio, participant demographics, human-robot roles, robot autonomy, input/output data, and equipment used. We propose respectively: a quantitative analysis allowing to observe objective metrics on trends and qualitative analysis of the relevant research topics to HRI covered by papers used in this review.

[1]http://www.aliz-e.org/
[2]http://dream2020.eu/
[3]http://chili.epfl.ch/cowriter
[4]http://www.squirrel-project.eu/
[5]http://www.l2tor.eu/

2 TECHNICAL OVERVIEW OF NAO OVER THE YEARS

NAO is 58 cm in height and weighs 5.6 kg. The robot is programmed by a specialised NAOqi framework, has an easy to use graphical programming tool Choregraphe (for complex applications and control of motions), and Monitor (for robot feedback and verification of joints or sensors), all of which allow to easily program and introduce the NAO behaviours (Bertacchini et al., 2017). It can be connected *via* wired or wireless (Wi-fi) network, thus allowing autonomous operation and remote control, which is important, especially when the robot is operating in a real-world setting. It has 25° of freedom, of which 12 for legs, five for the arms, two for the head, which enables it to move and perform actions. Furthermore, it has four directional microphones and speakers and two cameras that are necessary for basic modules such as built-in text-to-speech and speech recognition for 20 languages, object recognition, face detection, recognition, and tracking, all of which provide the possibility to act more naturally and human-like. **Table 1** presents an overview of NAO's hardware and software improvements over the years. For example, NAO's V3 in 2008 supported only nine languages, while the current V6 version provides support for 20 languages. Additionally, NAO's cameras, microphones, and storage were improved in three instances: from V3 to V4 or V5 to V6.

The first NAO driver for Robot Operating System (ROS) was released by Brown University's RLAB in November of 2009 (ROS, 2010) which supported head control, text-to-speech, basic navigation, and access to the cameras. Later, the University of Freiburg's Humanoid Robot Lab improved NAO's driver with new capabilities, such as torso odometry and joystick-based teleoperation. Already in December that year, the Humanoid Robot Lab released a complete ROS stack for the NAO that additionally contained IMU state, a URDF robot model, visualization of the robot state in rviz, and more (ROS, 2010).

Additionally, NAO users around the world had an opportunity to download an existing behavior or upload their

TABLE 1 | NAO's evolution in technical characteristics over the years.

NAO version	V3+ (2008)	V3.2 (2009)	V3.3 (2010)	V4 (2011)	V5 (2014)	V6 (2018)
Storage	2 GB Flash memory			2 GB+8 GB Micro SDHC		32 GB SSD
2 × Cameras	640 × 480, 30 fps 58 Diagonal Field Of View (47.8 Horizontal FOV, 36.8 Vertical FOV)			1280 × 960, 30 fps 72.6 Diagonal FOV (60.9 Horizontal FOV, 47.6 Vertical FOV)		640 × 480, 30 fps or 2560 × 1920, 1 fps 67.4 Diagonal FOV (56.3 Horizontal FOV, 43.7 Vertical FOV)
4 × Microphones	Sensitivity: −40 mV/Pa ± 3 dB Frequency range: 20 Hz–20 kHz				20 mV/Pa ± 3dB 150 kHz to 12 kHz	Omnidirectional 250 mV/Pa ± 3dB 100 Hz to 10 kHz
Languages	9 (English, French, Spanish, German, Italian, Japanese, Korean, Chinese, Portuguese)			19 languages (+ Arabic, Czech, Danish, Dutch, Brazilian, Greek, Polish, Finnish, Swedish, Russian, Turkish		20 languages (+ Norwegian)

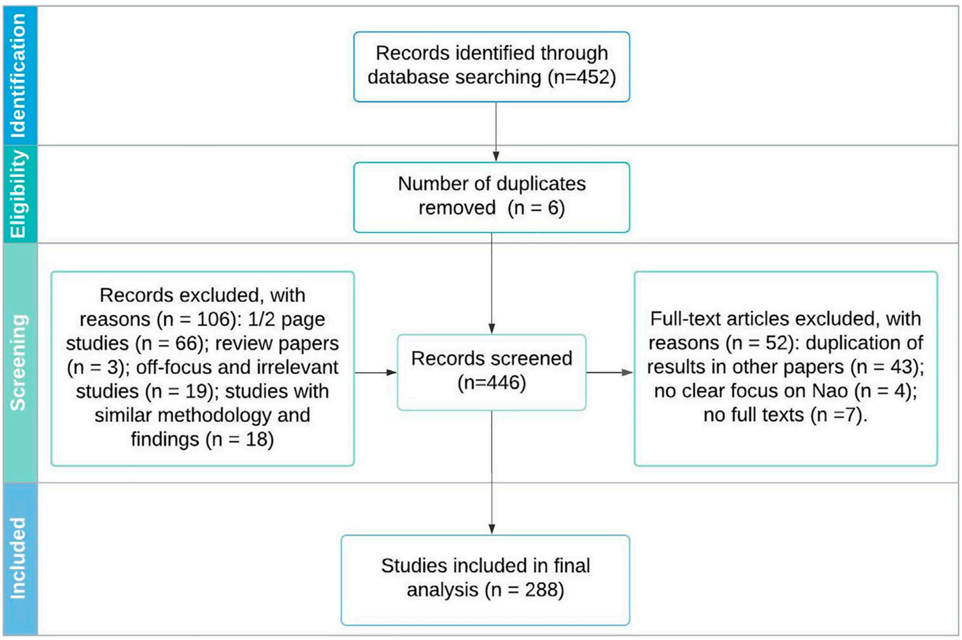


FIGURE 1 | The screening process (adapted from PRISMA template 2009).

own robot behavior to the Application Store. In 2014, ASK NAO[6] was released to support ready robot behaviors for conventional and special education. Similarly, but with a more general purpose, Zora Solution Software[7] was also offered to the market with more than 50 different robot activities to be used *via* a tablet by a non-technical user (such as a health professional).

3 METHODOLOGY

Our methodology followed similar works previously published in HRI and presenting a review of articles in the domain (Belpaeme et al., 2018; Johal, 2020; Obaid et al., 2020). We adopted a scoping review framework to extract relevant information from the literature to address our research questions. This approach is helpful to provide an overview of diverse research evidence in broad types of literature (Sucharew and Macaluso, 2019). We describe below the procedure carried out to collate the set of the relevant article and analyze their content in **Figure 1** which follow the PRISMA flowchart.

3.1 Identification

To identify potentially relevant documents, the Scopus[8] bibliographic database was searched for papers published

[6]https://www.asknao-tablet.com/en/home/

[7]https://www.robotlab.com/store/zora-robot-solution-for-healthcare

[8]https://www.scopus.com

from 2010 to October 2020. The term search was performed in October 2020. The Scopus database includes IEEE, Springer, and ACM DL and allows it to cover a wide range of publication venues. Because our goal is to broadly look at the research works done in HRI with NAO, we kept the search term open. We limited our search string to English-written publications as we searched for the terms "NAO" AND "human-robot interaction" in title, abstract, or keywords.

Overall, an initial 452 records were retrieved and underwent the screening process. They were stored on Zotero and then were exported into BibTeX and CSV. The following steps of the analysis of the collected documents were done by entering information on an online Google spreadsheet.

3.2 Screening Process

After identifying the records, we first consulted abstracts to ensure that they used NAO in the study. We excluded 106 studies provided only a quick overview (e.g., workshop, demonstration) in one or two pages in length. We removed the review and off-topic papers that lack any NAO intervention, both theoretically and practically.

In the second round, we consulted full texts to ensure that the chosen records do not replicate results. Since we had some studies produced by the same group of authors, we screened them in-depth and kept an extended version of the work. In addition, seven papers were excluded from review as we could not access full texts. As a result, we were left with 288 papers for the final analysis - annotation.

3.3 Coding Framework

To identify the categories for data analysis, we integrated and adapted the HRI taxonomies from previous studies (Yanco and Drury, 2004; Bethel and Murphy, 2010; Salter et al., 2010; Tsiakas et al., 2018; Baraka et al., 2020; Onnasch and Roesler, 2020) and annotated the papers by the predefined categories. We describe below the different annotations used. These were used to produce quantitative analysis and to identify trends.

3.3.1 Geographical Distribution

This information is not easy to infer from the publication; we chose to manually extract this information by checking the author's affiliation and address, and country on the paper. While not perfect, we believe that it should give us a reasonable estimation of the country where the research was conducted for most articles.

3.3.2 Type of Contribution

The field of HRI is very interdisciplinary. Inspired by the research themes of the ACM/IEEE HRI conference[9], we chose to annotate the type of contribution according to four themes:

- *User studies* provide rigorous data on and analysis of HRI in the laboratory or in-the-field settings. They also should present sound methodology (quantitative, qualitative, or both) and accurate analyses that result in novel insights and acknowledge the limitations and relevance of the methods. Papers that presented an empirical evaluation with human participants were annotated as a user study.
- *Technical* papers are motivated to improve robot's behaviors for the purposes of better interaction and collaboration with humans. The question of how technology advances HRI is key to these studies. They should include novel robot system algorithms, software development technologies, and computational advancements in support of HRI.
- *Design* contributions target research that takes a design-centric approach to HRI. They usually discuss the design of new robot morphologies and characteristics, behavior patterns, and interaction methods and scenarios, among many others. They should demonstrate essential or better interaction experiences or behaviors for robots.
- *Theory and methods* aim at unpacking fundamental HRI principles that include interaction patterns, theoretical concepts, updated interpretations of existing results, or new evaluation methodologies. Such papers might originate from original studies and existing research and methods or may take solely theoretical or philosophical perspectives.

3.3.3 Research Contributions

Looking at all the papers in the selection, we identified the main research objective (e.g., facial recognition, non-verbal communication, programming framework) for each paper. We then grouped these objectives into several classes of contributions: robot perception and recognition (emotion, facial, object, body, sound, speech, gesture, color, gender, text), robot's communication (verbal, non-verbal), reinforcement learning, and cognitive architecture. Imitation and display of emotions are separated from non-verbal communication due to a greater focus on them in observed studies. Apart from them, we included kinesthetic learning, physical exercises, taking an object, walking, and moving body parts. Some studies are both technical and user study, and there is more than one contribution example per paper.

3.3.4 Application Field

Baraka et al. (2020) provided a cross-sectional snapshot of key application areas for social robots, and, intuitively, robots are used in more than one field. Our categories included: autism therapy, education, elderly care, healthcare, learning disabilities, public service, entertainment, art, sport, and generic.

3.3.5 Human-Robot Ratio

Goodrich and Schultz (2007) considered that the ratio of people to robots directly influences the human-robot interaction. This taxonomy classification defines the number of a robot(s) and a participant(s).

[9]https://humanrobotinteraction.org/2021/full-papers/

TABLE 2 | The description of roles for participant and robot.

Role		Description
Participant	peer	interacts with a robot to achieve a shared goal
	coperator	works with a robot to fulfil a shared goal and does not directly depend on a robot
	collaborator	works as a teammate together for joint task completion
	learner	learns something from a robot
	imitator	imitates a robot's gestures or action
	interviewee	answers to the questions from a robot
	mentor	takes on a leadership or teaching role
	supervisor	monitors a robot and gives instructions on how to perform the task
	operator	is aware of where and what a robot is doing
	mechanic	works with robotic software or hardware and controls the physical setting
	information consumer	does not necessarily interact with a robot, but uses information that comes from it
	bystander	does not interact with a robot but shares the same space
Robot	peer	acts as a friend to achieve a common interaction goal
	learner	acquires new skills or behaviors from humans
	tutor	supports learning by being in a teaching position
	mediator	enables an interaction between two or more people, so that they can engage through a robot
	assistant	performs actions alongside humans (e.g. a teaching assistant)
	interviewer	asks questions
	demonstrator	shows model behaviors or actions
	testbed platform	validates or tests theories and algorithms in an experiment

TABLE 3 | The level of robot autonomy.

Level	Description
Wizard of Oz (Woz)	the robot is controlled by a human in the non-collocated environment where the robot is present
Autonomous	the robot acts based on its input without any external human control during decision-making
Combination	the robot integrates different levels of autonomy (e.g. controlled fixed command patterns)
Scripted/fixed	the robot follows scripted spatio-temporal command patterns, despite the external factors
Teleoperation	the robot is controlled by a human present in the same environment as the robot is

3.3.6 Participant's and Robot's Role

Goodrich and Schultz (2007) identified HRI roles, which were adopted by other researchers (Yanco and Drury, 2004; Tsiakas et al., 2018; Onnasch and Roesler, 2020). Based on their classification, 12 distinct participant's roles and eight robot's roles were defined. The description of each role is shown in **Table 2**.

3.3.7 Input and Output Channels

Onnasch and Roesler (2020) presented a taxonomy category which is named as the communication channels, split into input and output to highlight the human-robot interaction. Input describes how the robot "perceives" information coming from the human. Humans may provide information either using an electronic (e.g., remote control through the device), a mechanical (e.g., robot's kinematic movement), an acoustic (e.g., commands), or an optical channel (e.g., gesture control). In turn, the robot's output can be transmitted to humans through tactile communication (e.g., haptics), an acoustic (e.g., sounds), and a visual channel (e.g., eye movements). In the current study, the major distinction is that we view the input as any information coming from the environment (e.g., camera), while the output is what the robot produces through its channels (e.g., speech).

3.3.8 Robot's Autonomy Levels

According to Salter et al. (2010), the robot's level of autonomy is defined as shown in **Table 3**.

3.3.9 Experimental Methodology

Based on the classification proposed by Bethel and Murphy (2010), a study design is grouped into three categories:

- Within-subjects design - each participant undergoes the same experimental condition and is exposed to all levels of the independent variables.
- Between-subjects design - participants are exposed to different groups where each group experiences different conditions.
- Mixed-model factorial design - the use of both between-subjects and within-subjects design components.

3.3.10 Duration of Interaction

Human-robot interaction studies can be grouped on the basis of the duration of interaction, which means the certain period of time when the human interacts with the robot (Baraka et al., 2020). Albeit it is challenging to define set boundaries between interaction times, we decided to follow the proposed duration

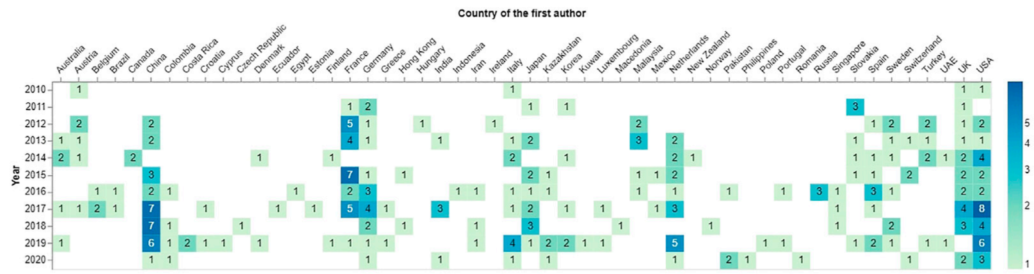


FIGURE 2 | Number of publication per country.

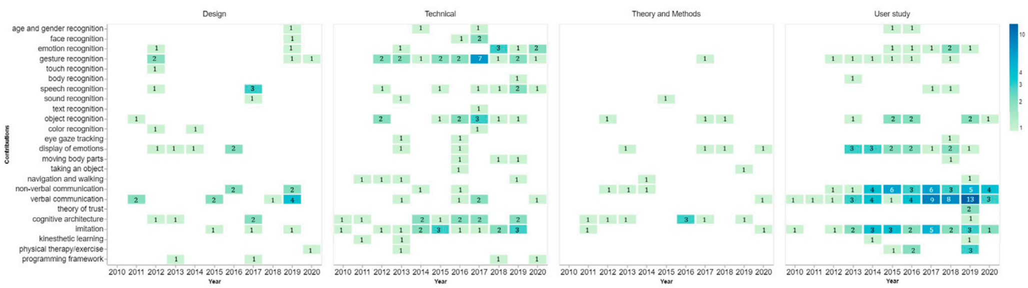


FIGURE 3 | Contributions made over the years grouped by each study type.

looking at the number of sessions. We annotated according to the following categories: short-term (single or few interactions), medium-term (several days or weeks), long-term (extended period).

4 QUANTITATIVE RESULTS

We propose to address our research questions with quantitative analysis to look at research trends over the years and the different categories identified above. All the graphs were generated using Altair, which is the declarative statistical visualization library for Python (VanderPlas et al., 2018).

4.1 Geographical Distribution

Figure 2 shows the frequency of publications across countries and per year. Earlier works that date back to 2010 were produced in anglophone countries such as the US and UK and European countries including Austria and Italy. France being the NAO's homeland, it also figures among the countries reporting a lot of research works. From the figure, it is apparent that the (predominantly) English-speaking world continues to dominate the HRI research with NAO. When compared to other parts of Europe, Nordic countries and Eastern Europe are substantially underrepresented. Notably, NAO has been used regularly in economically wealthy Asian countries such as China and Japan. Over the years, the largest number of papers were published by researchers from the USA ($N = 33$), China ($N = 30$), and France ($N = 25$). These results may serve as an example of widening digital inequity between countries with different economies.

Having said that, it is interesting to note that NAO was used quite broadly around the globe. Evidently, increasing the number of languages supported by the robot as shown in **Table 1** has been an important factor in the integration of the robot. The language options for its text-to-speech API covering 20 languages can explain this broad use. We also can note that this multilingualism supports cross-cultural reproductibilty of previous studies and theories that were tested with NAO.

4.2 Research Contributions

Figure 3 demonstrates research topics that were identified as the papers' main contributions. We group them by paper type and show their frequencies over the years. As of 2010, earlier contributions represent verbal communication, cognitive architecture, and imitation in technical and user studies. We cannot observe straightforward trends in design, theory and methods, but verbal communication and cognitive architecture seem to have a proper representation as common contribution topics. Our analysis shows that verbal (e.g., dialogues) and non-verbal communication (e.g., joint attention) were the most common contributions among user studies published in 2019. Gesture recognition was generally observed to be a popular contribution topic in technical papers, especially in 2017. Color, face, touch, and sound recognition were among the least popular topics for contributions, probably because of NAO's limited perception abilities. It is important to note that some technical contributions (e.g., emotion recognition) are

FIGURE 4 | Application fields grouped by each study type over the years.

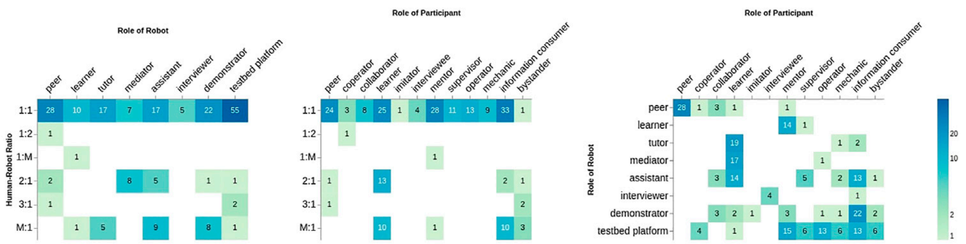


FIGURE 5 | Human-robot ratio per role of robot and participant, and co-occurrence of robot and participant role **(right)**.

present within user studies or theory and method groups due to a single paper having several contributions. The more consistent distribution of design, theory and methods, and technical contributions, and the increasing rate of user studies through the years shows how the first three have contributed to the integration and testing of the robot in various domains through user studies.

4.3 Application Fields

The applications contexts of NAO are displayed in **Figure 4**. Evidently, generic fields are prevalent across all types of studies. This hints on how the community has been keen on developing the generic technology with the NAO robot with the goal of integrating it in various applications. Which, in turn can contribute to integrating the robot not only in the research domain but also in the real-world applications. Furthermore, this means that NAO is being used for non-specific purposes such as addressing general hypotheses and technical questions, as can be seen from the share of technical studies. In user studies, the use of NAO has expanded stably in healthcare, autism therapy, and education since 2015. We separated studies on autism therapy from healthcare as this context is receiving a growing attention within HRI. Some unusual application areas are space (helping astronauts in space flight-operations in Sorce et al. (2015), art (drawing on canvas in Gurpinar et al. (2012) and performing in theatrical play in Petrović et al. (2019).

4.4 Human-Robot Ratio

Figure 5 displays the ratio of participants to robots for various kinds of robot and participants' roles. The vast majority of studies used one-to-one interaction with the roles of the robot as a testbed platform (N = 55) and the role of the human as an information consumer (N = 33). In a dyadic interaction, the robot quite often played a role of a peer (N = 28), demonstrator (N = 22), tutor (N = 17), assistant (N = 17), followed by learner (N = 10), mediator (N = 7) and an interviewer (N = 5). Participants often played the role of a mentor (N = 28), learner (N = 25), and peer (N = 24).

The ratio of many participants to a robot (M: 1) comes second with the robot roles of assistant (N = 9) and demonstrator (N = 8). In this context, humans were introduced as information consumers and learners in 10 studies for each. Triadic interaction was common among mediator and assistant robot roles and human learners (N = 13). Only a few studies had the ratio of 3 : 1 with no obvious trends.

The first trend shows that the majority of studies were carried out using dyadic interactions. The limited number of studies with two robots or more can imply either on the difficulties of developing multi-robot interactions or lack of interest in the community. Furthermore, while there are quite a few number of studies on triadic interactions with two humans and one robot, they are still limited to specific types of interaction where the human is a learner or an information consumer. On the other hand, after dyadic interactions, the most number of publications were carried out with one robot to more than five human ratio, with the robot being a demonstrator, assistant, or tutor. The analyses shows the number of studies using such dynamic has increased over the years.

4.5 Human-Robot Roles

In **Figure 5** (right), we also demonstrate robot-participant interaction roles. It becomes clear that NAO generally plays collaborative and mediating roles. Our analysis shows that the most common HRI roles with NAO have been: peer-to-peer (N = 28) and demonstrator-to-information consumer (N = 22). When the human was in the learner's role, a robot was most frequently in the role of either a tutor (N = 19), mediator (N = 17) or an assistant (N = 14). Our analysis presents the interviewee-

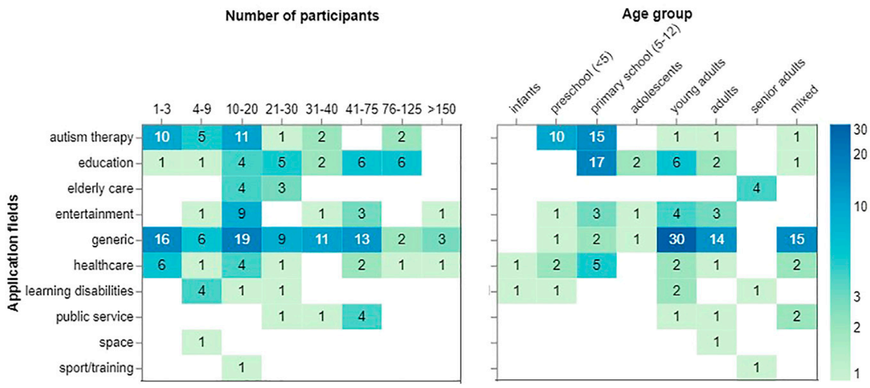


FIGURE 6 | Number and age group of participants per application field.

interviewer dyad as an exceptional and interesting case in HRI. The examples of peer interaction include learning a language (Kim et al., 2019), playing a game (Hirose et al., 2015), and working as a team (Mubin et al., 2014). NAO as demonstrator or presenter performs actions in front of the participants (Krogsager et al., 2014; Wang et al., 2014; Kaptein et al., 2017). Learner-tutor interaction occurs in educational settings where NAO is engaged with language teaching (Kose et al., 2014; de Wit et al., 2020) and providing emotional support (Miskam et al., 2015; Cuadrado et al., 2016). NAO as a mediator predominantly helps children with autism or learning disabilities to scaffold social and communication skills (Shamsuddin et al., 2012; Huskens et al., 2016; Ioannou and Andreva, 2019). A further analyses of the dynamics between the role of the robot versus participant show some pairs of roles appear more than others. For example, there is a consistent use of robot as a testbed platform with human acquiring various roles such as mentor, mechanic, or information consumer. On the other hand, we can see lots of studies with human as a learner where the robot might have been a tutor, mediator, or assistant. It is also important to mention, some dynamics such as peer/peer or learner/tutor are more common in education and autism therapy.

4.6 Number of Participants and Age Groups

Figure 6 juxtaposes how often the user studies had various ranges of participants. For the most part, the number of participants ranges from 1 to 20 people, having the greatest number in the range "10–20." A smaller number of participants (up to three people) is mostly used for autism therapy, generic, and healthcare applications. A fair amount of generic studies recruited a large number of participants ranging from 30 to 75. Interestingly, studies conducted for education recruited the biggest number of participants that can go up to 125 people. There were a few entertainment, generic, and healthcare studies that had more than 150 participants.

Figure 6 (right) demonstrates the total number of studies that had various age groups for each application field. Children at preschools and primary schools participate in studies that focus on education ($N = 17$) and autism therapy ($N = 25$). Generic fields work with much older age groups since the studies are typically conducted with university students or staff (e.g., (Stadler et al., 2014; Celiktutan and Gunes, 2015)). The figure also reveals that senior adults interact with NAO for elderly care and learning disabilities applications. Infants and adolescents are the least represented age groups.

Figure 6 (left) shows that some application types such as autism therapy and healthcare use a smaller number of participants per study (< 20). A quick look at the distribution of age groups in autism therapy showed more focus on preschool and primary school aged children. This can explain the possible difficulties in recruiting participants for autism therapy studies which can be one of the causes of small sample sizes. On the other hand, educational user studies tend to have a higher number of participants (between 20 and 125) with the age group distribution of primary school and young adults. One of the interesting trends is the higher population of young adults and adults in generic studies, which can be explained by the possible easier procedure to recruit them for user studies. Whereas, most studies with children and the elderly that might be harder to recruit are conducted for specific applications such as autism therapy, education, and elderly care.

4.7 Input and Output Data

Figure 7 provides the frequency of input and output data throughout the application fields. Primarily, generic studies deployed speech (N = 36), full-body (N = 27), face (N = 22), and gestures (N = 21) as an input data for recognition. Interestingly, tactile input is mostly used within generic types of applications, with a few studies in autism therapy, elderly care, and learning disabilities. Tablet and mobile devices were mostly used for autism therapy, education, and generic fields. The least popular types of input data come from wristbands and electroencephalography (EEG). This might be due to the intrusive features of most wearables.

In line with these results, NAO's output data is mostly speech and gestures in generic fields, autism therapy, and education. Eye contact and LEDs were comparatively less used by the robot.

Considering the various types of studies conducted with the NAO robot, we also looked at the type of equipment used alongside the robot. **Figure 7** shows the input data (left),

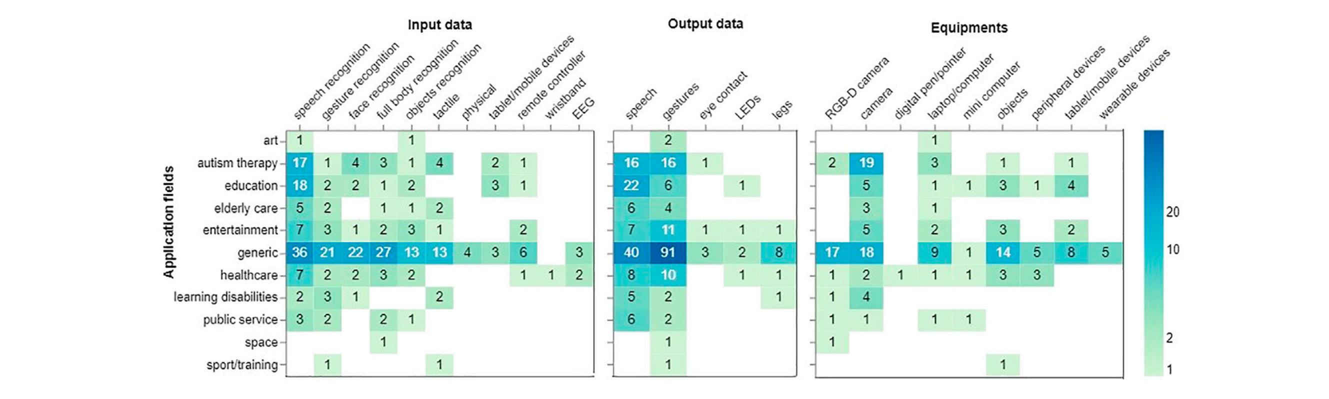


FIGURE 7 | Input data **(left),** output data **(middle),** equipment **(right)** for each application field.

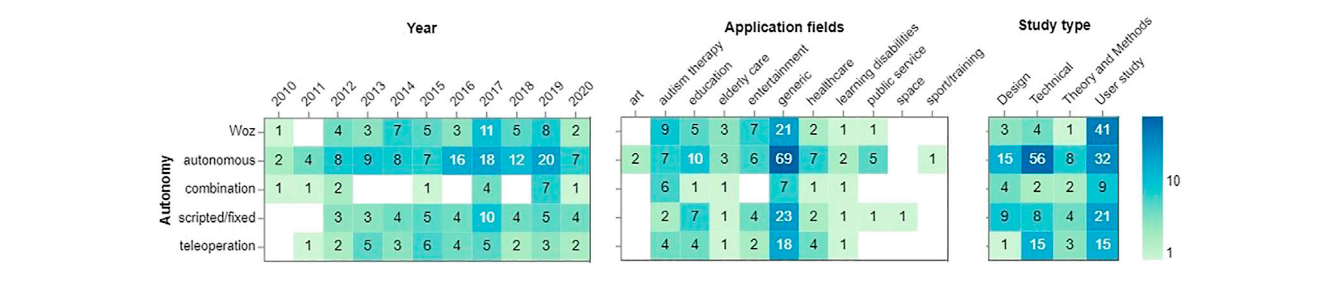


FIGURE 8 | Robot's autonomy per year and application fields.

output data (middle), and equipment (right) used over all application fields. Speech recognition dominates the type of input data, which has been used on almost all of the application types, and it is proceeded by gesture, face, body, and object recognition. It is notable that apart from generic applications, higher use of speech recognition can be seen in autism therapy and education. Considering the target age groups for these applications, this calls for more attention in developing speech recognition technologies for children. As for the output data, 7 (middle), most applications seem to have utilized the robot's speech and gestures. Autism therapy, entertainment, and healthcare had shown a higher tendency of using gestures in comparison to other applications.

4.8 Equipment

Figure 7 (right) also presents the use of different equipment that researchers make use of during their user studies. The most popular equipment are RGB-D cameras, ordinary cameras, and physical objects (e.g., a ball, geometric figures). Again generic studies employed these equipment more often than any other field. Tablet or mobile devices are commonly used in educational settings. Some examples of wearable devices are a helmet with electrodes (Gomilko et al., 2016), pluggable eyebrows to express emotions (De Beir et al., 2016) and peripheral devices such as a microphone, keyboard, and LCD monitor to show visual stimuli (Hu et al., 2016). Looking at the additional equipment used with the NAO robot, one notable trend is the additional usage of the camera and RGB-D camera alongside the NAO robot. While the camera might have been used to provide additional data from different angles to analyze the participant or the interaction, the use of RGB-D cameras, specifically in case of its placement from the robot's point of view, can hint on the possible use cases of adding such a gadget to the robot, even as a supplementary item. Other equipment frequently used are laptop/computer and objects which depending on the activity, can add more interaction dimensions and modalities to the robot.

4.9 Robot's Autonomy

Figure 8 illustrates the levels of robot autonomy by year and application fields. We observe clear trends that NAO is becoming more autonomous in recent years, with a significant increase from 2016 to 2019. Wizard of Oz is the second most widely chosen control type that has been evenly spread across the given years, except for 2011. Only generic fields appear to use all levels of robot autonomy, as a rule, autonomous mode, when compared to other fields. Novel application fields (space, art, and sports) constitute the least share in robot autonomy. Essentially, we can also report that technical studies use autonomous mode, while user studies give preference to the WoZ setting. In fact, a robot's autonomy greatly varies in user studies as the modes are divided proportionately. The combination mode appears to be unpopular across all study types.

As we know NAO robot comes with NAOqi, a visual interface called Choregraphe and can be programmed using

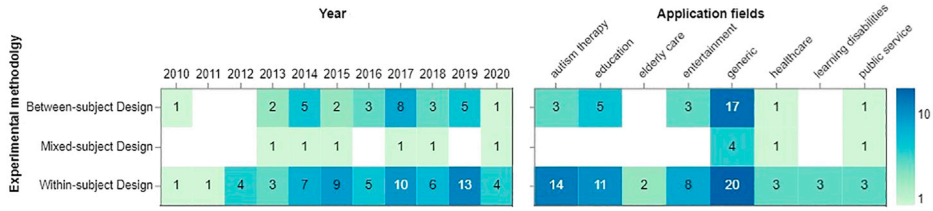


FIGURE 9 | Experimental methodology per year **(left)** and application fields **(right).**

ROS. These all give the user plenty of opportunities to develop behaviors and interactions with the robot. As a result, in **Figure 8**, we looked at the distribution of the robot's autonomy over the years (left), per application fields (middle), and study types (right). One noteworthy trend is the increasing rate of studies with the fully autonomous robot through the years, more specifically in 2016, 2017, and 2019. This can hint on how the technical developments and increasing interest in using the robot have contributed to the more autonomous deployment of the robot. After generic applications, education, autism therapy, and healthcare had the highest population in using NAO robot autonomously. It is worth mentioning that more studies in autism therapy have used Wizard of Oz than fully autonomous, which can also be explained by the restriction associated with running studies in this field. Looking at the autonomy versus study types (right), it can be seen that Wizard of Oz autonomy was more popular in user studies which can be explained by considering the difficulties of deploying a fully autonomous robot to interact with users. On the other hand, the fully autonomous robot has been used more in technical studies, then in user studies, and finally in design studies.

4.10 Experimental Methodology

Figure 9 illustrates the frequency of using three types of experimental methodology across years and application fields. Seemingly, a within-subject design was commonplace from 2012 onwards. It reached the maximum number of publications (N = 13) in 2019, and the three most common fields of its use are generic, autism therapy, and education. Generic fields again lead the field by deploying both within- and between-subject design. Studies on autism therapy and education adopt the two designs. Studies in healthcare and public service choose between-subjects design rarely than any other field.

We have also analyzed the experimental methodologies used in user studies, both through the years and based on application fields as shown in **Figure 9**. As seen from the figure, the use of within-subject experimental design has increased through the years, and it is generally used more than between-subject and mixed designs. And among application fields, autism therapy, education, and entertainment were more prone to using within-subject designs. Apart from methodology, we also looked at experiment duration, as categorised in short, medium, and long-terms.

4.11 Duration of Experiment and Sessions

Figure 10 shows how long human-robot interaction lasts across years and fields. We see clearly that the majority of studies are short-term, specifically between 2015 and 2019. This obvious trend is explained by the prevalence of generic fields. Medium-term and long-term studies were scarce before 2014, but their numbers relatively increased by the mid-2010s. Only several studies that focus on autism therapy with NAO used a long-term approach. Despite no explicit trends, we can observe that the interaction up to 30 min is more common compared to other periods of time, mostly in generic and autism studies. Considerably, a few studies (N = 5) lasted for more than 60 min.

Figure 10 (left) shows the majority of the studies have been conducted on a short-term basis, and as the number of studies increased through the years the number of short-term studies has increased as well. There is no visible trend of increasing long-term studies at least with the NAO robot which can be thought provoking and worth understanding its underlying causes. As human-robot interaction field is thriving to understand the dynamics between human and the robot, we need more long-term studies to be able to show how the robots can integrate into our lives and society. Looking at **Figure 10** (middle), we can see all generic studies have been conducted with short-term duration. It is intuitive to conduct a short-term study when developing or testing technology for generic purposes and invest more in running long-term studies with the specific application in mind. For example, studies on autism therapy and healthcare were more likely to have medium and long-term duration than the rest of the applications. The **Figure 10** (right) shows a quick overview of the duration of the sessions in minutes. The duration of sessions is a function of the application and the interaction; hence we cannot observe a particular trend. However, it is interesting to see that people have been participating in experiments with NAO robots that had lasted up to 120 min. In general, the more we are trying to integrate robots into society and move them from research labs into the wild, we might need to run more long-term studies.

4.12 Concluding Remarks

The noteworthy findings that emerge from this quantitative data are:

- While studies with NAO have been produced all over the world, the great majority of studies are published by

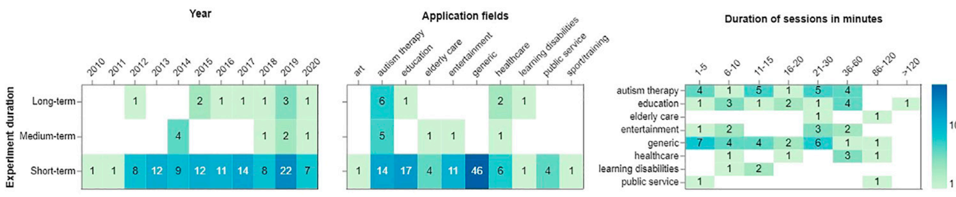


FIGURE 10 | Timespan of the experiment per year **(left)** and application fields **(middle)** and duration of sessions in application field **(right).**

researchers in the Global North, particularly in the U.S and Western Europe.

- NAO has been used for generic purposes, yet it appears to gain traction in autism studies and education since 2015.
- Despite physical limitations, speech, and gestures are the main communication channels for NAO to interact with the environment. The lack of accurate speech recognition and natural behaviours such as emotions and spontaneity causes mixed feelings about its social use.
- Although efforts have been made to allow NAO to function autonomously in generic fields, it still depends on human control and supervision when interacting with the end-users (such as children).
- Humans from different age groups can interact with NAO, depending on the variation in contexts of use. Therapeutic and educational studies recruit primary age children, while generic studies mix up all available age groups. Dyadic interaction prevails significantly.
- The most recurrent robot roles for NAO are found to be peer, demonstrator, tutor, and mediator.
- The studies with NAO are predominantly short-term and may last for approximately 15–30 min.
- The available studies apply within-subject design more often than between-subject or mixed-subject. This is indicative of its relatively easier process as the number of participants can be smaller.

5 QUALITATIVE RESULTS

We also conducted a qualitative narrative review to discuss the primary research focus reported in the papers present in our collection. This section is concluded with the key findings that emerge from the literature.

5.1 The Human Perception of NAO

The way robots interact with humans is critical to evaluate the overall quality of robot capabilities. HRI has drawn great attention in studying how humans perceive robots in terms of their appearance, task performance, and communication skills, among many other robot features. User perceptions and experiences with NAO vary from one context to another as well as between user populations, including children, parents, teachers, and experts. Due to its small size, NAO is predominantly used in the child-robot interaction scenarios (*see* **Figure 6**), with some exceptions, in elderly care. Nevertheless, the majority of users perceive NAO as a friendly and sociable robot (Hernandez-Cedeño et al., 2019; Turp et al., 2019). There were also reports of mixed feelings about the robot, considering its physical and technical limitations (Cruz Maya et al., 2015; Sarabia et al., 2018). Additionally, the human-like appearance and non-judgemental characteristics of NAO are highly appreciated by users (Henkel et al., 2019; Olde Keizer et al., 2019). Users would like NAO to be more emotionally expressive, responsive, and have a natural voice and gesturing (Anastasiou et al., 2013; Ahmad et al., 2017). Authors used a variety of questionnaires to evaluate NAO's characteristics and performance based on: anthropomorphism (Zlotowski et al., 2014; Kraus et al., 2016), user experience (Alenljung et al., 2018; Olde Keizer et al., 2019), user acceptability (Ahmad et al., 2017), robot personality (Liles and Beer, 2015; Peters et al., 2017; Kraus et al., 2018), robot behaviors (Pan et al., 2013; Njeri et al., 2016; Rossi et al., 2019), user expectations and evaluation (Anastasiou et al., 2013; Henkel et al., 2019), and perceived trustworthiness (Jessup et al., 2019). **Table 4** presents common questionnaires that are used in evaluating human-oriented perception of NAO.

When touching the robot, positive experiences with NAO were characterized as fun and engaging, while negative experiences were described to be odd and unsafe due to its small size and hard surface (Alenljung et al., 2018). Comparing low and high interactivity, Tozadore et al. (2017) found that children enjoy their experience with the high interactive NAO that use a warm greeting and recognizes their names. When compared to the virtual agent, users still favored NAO to be engaging and responsive (Artstein et al., 2017). Both teachers and students were comfortable with the use of NAO, yet they emphasised the need for facial expressions in NAO (Ahmad et al., 2017). Gender might influence how robots are perceived. For example, users found a male NAO more trustworthy and competent than a female one, which was only rated as likable (Kraus et al., 2018). In another study, children at different developmental stages had varying preferences towards NAO's gender: younger children (5–8 years old) wanted a robot that matched their own gender, while older children (9–12 years old) did not have such gender-driven preferences (Sandygulova and O'Hare, 2015).

TABLE 4 | Perception questionnaires commonly utilized in the reviewed studies.

Name	Author	Measurements	Item type
Godspeed Questionnaire Series (GQS)	Sturgeon et al. (2019)	anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety	5-item with 5-point Likert scales
Unified Theory of Acceptance and Use of Technology (UTAUT)	Sinnema and Alimardani. (2019)	anxiety, attitude towards technology, perceived enjoyment, perceived sociability, perceived usefulness, social influence, and trust	7-item with 5-point Likert scales
Negative Attitude Toward Robots Scale (NARS)	Mirnig et al. (2017)	attitude toward interaction with robots, social influence of robots, and emotions in interaction with robots (e.g. I would feel relaxed talking with robots)	10-item with 5-point Likert scales
System Usability Scale (SUS)	Olde Keizer et al. (2019)	attitude towards usability (e.g. “I thought the system was easy to use,” “I felt very confident using the system”)	10-item with 5-point Likert scales
Individual Differences in Anthropomorphism Questionnaire (IDAQ)	Zlotowski et al. (2014)	anthropomorphic (“durable,” “useful,” “good-looking,” “active” and “lethargic”) and nonanthropomorphic traits (intentions, emotions, consciousness, free will, mind)	30-item with 10-point Likert scales
Complacency-Potential Rating Scale (CPRS)	Zlotowski et al. (2014)	attitudes towards automation (confidence-related, reliance-related, trust-related, and safety-related complacency	20-item with 5-point Likert scales
Propensity to Trust Technology (PTT)	Jessup et al. (2019)	attitudes towards technology and collaboration with technology (e.g. “Generally, I trust technology”; “Technology helps me solve many problems”)	6-item 5-point Likert scales
Robot Interactive Experiences Questionnaire	Mubin et al. (2014)	attitudes towards engagement and social interaction (e.g. alive, friendly, social)	8-item with 7-point Likert scales

5.2 Verbal Communication With NAO

NAO can also be presented as a conversational companion to assist people with daily tasks, serving as language tutors in autism therapy (Fuglerud and Solheim, 2018), facilitators in speech therapy for hearing-impaired children (Ioannou and Andreva, 2019), and peers for self-disclosure among people with visual impairments and intellectual disabilities (Eyssel et al., 2017; Groot et al., 2019). Interestingly, NAO can also act as a storyteller that sustains children’s attention due to novelty and gesture frequency, while human storytellers may become fatigued to deliver the story (Wu et al., 2018; Ruffin et al., 2020). These studies also suggest that timely feedback during human-robot communication has been regarded as a success factor contributing to the quality of interaction. By presenting NAO as a talking partner, Omokawa et al. (2019) distinguished between two dialog types for verbal interaction: query type is a question-and-answer format, and phatic type is a casual format that involves small talk and/or personal feelings (e.g. acceptance). As the authors noted, human utterances are hardly recognized in the latter dialog type due to short words that probably express emotions. In Da Silva et al. (2018), NAO as a motivational interviewer also enabled verbal communication with humans, yet its lack of personalization was disliked by many participants (e.g. repeating the question an user had already answered). Recently, Graña and Triguero (2019) proposed a spoken dialogue system for the NAO to learn to answer autonomously based on human input[10]. For human-friendly communication, Manome et al. (2020) developed a machine translation system in which the NAO correctly speaks Japanese words that were converted into morphemes to enable easier pronunciation. The examples above indicate great prospects for the NAO to improve its verbal skills that are necessary for natural communication with humans.

[10]https://zenodo.org/record/2567595#.YPVq0y8RoUs

5.3 Non-verbal Communication With NAO

In the same way, NAO’s non-verbal social cues play an important role during human-robot interaction. Non-verbal communication happens in many instances that help facilitate joint attention, turn-taking, shared attention during HRI. Although NAO lacks orientable eyes, which may be counted as a serious limitation, results indicate that head rotations typically help imitate eye contact (Cuijpers and van der Pol, 2013). For instance, NAO can serve the needs of children with autism who often find eye contact with other people uncomfortable and therefore try to avoid it. Additionally, different visual stimuli such as changing eye colour cyclically and blink by NAO were added to encourage eye contact with children (Ismail et al., 2012; Ali et al., 2020). Eye contact and turn-taking usually fit together, for example, when children kick the ball and wait for NAO to kick it back (Tariq et al., 2016). Gazing behavior, however, is the important sign of communication because it allows people to infer engagement and intent. NAO gazes over objects of its attention and points to them to elicit joint attention (Anzalone et al., 2015). These examples demonstrate the extent to which the child’s eye movements would be responsive when NAO directs its attention to other objects. NAO was able to perform Turkish Sign Language gestures (Kose et al., 2014). In a buyer-seller negotiating, a human-robot handshake prior to negotiation may benefit both sides to reach a more positive outcome (Bevan and Stanton Fraser, 2015).

5.4 NAO as a Support for Affective Computing Research

NAO cannot directly express emotions through facial expressions, yet it can perform acoustic and physical expression of emotions. It is viewed as one of the limitations in its design. Most research studies proposed to express emotions through utterances (De Beir et al., 2016), gestures (Beck et al., 2012; Erden, 2013; Miskam et al., 2013; Rudovic et al., 2017), or both (Aly and Tapus, 2013; Tielman et al., 2014; Miskam et al., 2015). A few others attempted to use innovative ways such as eyebrows showing emotions (De Beir et al., 2016) and motion planning for four emotional patterns (Wei and Zhao, 2016).

Beck et al. (2012) designed six key poses that were implemented on the NAO to display emotions such as anger, pride, sadness, happiness, fear, and excitement, as captured by a motion camera system. Similarly, Erden (2013) adapted emotional human postures to the robot that expressed anger, happiness, and sadness through 32 different postures for each emotion. Creatively, De Beir et al. (2016) used 3D-printed and wearable eyebrows that allow NAO to show anger or sadness while doing other tasks simultaneously. In a play scenario, NAO can also show excitement and enjoyment using matching phrases such as "I am really excited!," "I enjoy playing with you!" while no emotional posture or saying is expressed in a boring or tiresome state (Franco, 2015). Miskam et al. (2015) proposed to use NAO for teaching emotions using LEDs, hand or body gestures to the children with autism, who then imitate the robot by repeating the emotions such as being happy or hungry. Rudovic et al. (2017) also used NAO for robot-assisted autism therapy, where children had to recognize different robot emotions as shown in emotion cards. Through Laban Movement Analysis (LMA), Wei and Zhao 2016) integrated four emotional patterns into robot behaviours using motion planning. Interestingly, (Manohar and Crandall, 2014) studied how novice people program robot's behaviors to express emotions through recorded audios and gestures and then recognized them. The study found that non-verbal emotions were not easy to discern than those expressed *via* verbal channels. NAO can also recognize human emotions through speech cues and facial expressions. For instance, Bechade et al. (2015) proposed an emotion recognition game in which the robot had to recognize emotions through the speech of humans. Likewise, Diaz et al. (2018) implemented a text2emotion system that enables NAO to execute behaviors based on its ability to recognize audiovisual stimuli. Stojanovska et al. (2018) tested NAO's emotion recognition rate by recruiting participants to act emotions in front of the robot. Lopez-Rincon (2019) enabled NAO to recognize human emotions based on their photos on the computer screen, from which NAO detected one-half of the face images (535/1192) from the Child Affective Facial Expression (CAFE) dataset. Roshdy et al. (2019) applied a human brain-based mapping system for emotion recognition through Emotiv headset, motivated by the mapping of the human brain activity into NAO. In general, the robot can express and recognize emotions successfully except if users are not good at displaying them.

5.5 NAO as a Tool for Therapy and Learning

Despite many generic use cases implemented for NAO, this robot is widely deployed within medical and educational institutions for use by children.

Learning by imitation refers to observing and performing a new behaviour by replicating the action of others. Within HRI, imitation is generally practiced with children with health problems (e.g., autism) because they have difficulties in motor and/or turn-taking skills. When children mirror robot gestures and other behaviours, they can improve social interaction skills. In this way, most human-robot interaction occurs in a playful environment, where children, robot, or both imitate. In Arent and Kruk-Lasocka (2019), NAO played two interactive games with children to improve their turn-taking skills through movement imitation. Arias-Aguilar et al. (2017) designed a child-robot interaction in which NAO proposes typically developing children to a "play" by imitating the same arms and legs movements that it makes itself. Chevalier et al. (2017) designed a playful task in which NAO performed several hand gestures in the background of music with the same duration and rhythm. Both the robot and children with ASD had to imitate each other's arm movements, but children were a bit confused to initiate them. In Di Nuovo et al. (2020), NAO presented itself and engaged with the children by playing music and storytelling and then asked to imitate its dance movements. Greczek et al. (2014) developed a Copy-cat game played between an NAO robot and a child with ASD. In the game, the robot asks a child to mirror its pose, saying, "Can you copy me?." In learning by imitation framework, some authors propose to use Dynamic Time Warping that observes joint angles trajectories instead of Hidden Markov Models (HMM) for time normalization (Thobbi and Sheng, 2010).

Meanwhile, some researchers (Cazzato et al., 2019) proposed a system where NAO recognizes the presence of a user in real-time and imitates the human's head pose. To augment motor skills, NAO may encourage imitation learning (e.g., sit-to-stand) in children with cerebral palsy, despite its physical limitation to move naturally (Rahman et al., 2015). In Ros et al. (2014), NAO taught children dance moves while providing verbal support with music. Tapus et al. (2012) developed a motor imitation task in which NAO imitates gross arm movements of children with ASD in real-time. The results show a high variation in children's reactions to the NAO, which means that not all children can benefit in the same way. For rehabilitation and prevention of scoliosis (abnormal curve of the backbone), Vircikova and Sincak (2013) presented NAO in hospital and school settings. The participating children imitated NAO's motions accurately, which also increased their motivation to exercise more. Quite similarly, NAO, as a trainer, performed physical exercises with elderly people who tried to imitate movements (Werner et al., 2013). In this context, users had to imitate mood-modified NAO's arm gestures in a game, after which the robot provided verbal feedback about user performance (e.g., "Yes, those were the right gestures" for a correct movement). Imitation is one of the important skills for individuals with developmental disorders who need to understand social cues from a young age. Therefore, research shows that NAO is able to facilitate initiation and turn-taking skills through imitation tasks or games.

NAO is generally welcomed by students who view this robot as a learning peer, a more knowledgeable tutor, or a less knowledgeable learner (Johal, 2020). Rosenberg-Kima et al. (2019) found that the physical presence of robots brought positive changes for university students because of the technical functionality, social, and psychological activity. Namely, students pointed out the benefits as follows:

"accessible to multiple people," "immediate feedback," "he is not judgmental like human beings," "pleasant and motivating." Some research has targeted specific skills required for language learning: reading (Yadollahi et al., 2018), grammar (Belpaeme et al., 2018), handwriting (Hood et al., 2015), alphabet (Sandygulova et al., 2020) or vocabulary learning (Balkibekov et al., 2016). Other research demonstrated that learners cultivate favorable impressions toward robots as learning companions, and the child-robot interaction may lead to increased self-confidence (Hood et al., 2015) and better task performance requiring creativity and problem-solving. Other studies e.g., Vogt et al. (2019) explored long-term learning between NAO and children to better understand this type of HRI in a real-world environment.

5.6 Typical Comparisons in HRI Studies With NAO

To identify the robustness and applicability of the social robot, comparative studies have been undertaken in terms of interaction roles and behaviors. Comparison occurs not only between robots but also between participants and interaction types. The comparisons between humans include children vs. adults (Kaptein et al., 2017), expert vs. non-expert (Ansermin et al., 2017), autistic vs. typically developing children (Anzalone et al., 2015), programmer vs. non-programmer (Stadler et al., 2014), and people from different cultures (Rudovic et al., 2017; Shidujaman and Mi, 2018). This shows that different groups of humans may have different experiences with a social robot.

Bethel et al. (2013) compared a human interviewer and a robot interviewer to find out which of them impacts participants when presented misleading information. The results show that the misinformation effect was significant in the human interviewer condition than in the robot interviewer condition. The authors suggest that its TTS system caused the lack of speech comprehension, which results in issues with the robot's understandability. In Henkel et al. (2019), participants found the robot interviewer as non-judgemental with whom they were likely to share secrets. In a language learning context, a human teacher and robot teacher performed sign language gestures in a real and virtual environment, addressing the embodiment effect (Kose et al., 2012). Tapus et al. (2012) explored whether children with autism engage more with a robot partner or a human partner during a movement imitation task, in which no significant differences were found. In performing physical exercises (Werner et al., 2013), users perceived NAO as less motivating than humans, but they also rated the robot as more motivating than a standard training plan they use regularly.

When exploring robot embodiment, most users perceive NAO better in terms of its engagement and social characteristics. Artstein et al. (2017) found that a physical robot was more preferred and engaging to participants when compared with a virtual agent, which in turn led to better memorization over a longer period. Bevan and Stanton Fraser (2015) were interested in comparing telepresent NAO against non-telepresent NAO when shaking hands with participants during negotiations, whereas Tozadore et al. (2017) evaluated controlling the robot autonomously and through WoZ. Both studies suggest that a robot's presence did not affect the degree of trustworthiness and appraisal, and user enjoyment, but the perceived level of robot intelligence may decrease when people know about teleoperation. Some studies explored robot personality effect on interaction quality such as extroverted vs. introverted (Aly and Tapus, 2013; Celiktutan and Gunes, 2015), low interactive vs. high interactive (Tozadore et al., 2016; Horstmann and Krämer, 2020), active vs. passive (Mubin et al., 2014), affective vs. non-affective (Tielman et al., 2014), emotional vs. unemotional and high vs. low intelligence (Zlotowski et al., 2014), lack of ability vs. lack of effort (van der Woerdt and Haselager, 2019), and simulated vs. real robot (Riccio et al., 2016). The robot-to-robot interaction and comparisons were also carried out in different contexts. However, only some papers compared the efficacy and utility benefits of the robots, mainly using the other robot as an alternative to the NAO or vice versa. Although children prefer NAO, they find easier to understand the gestures of a taller R3 (Kose et al., 2014) and rate Baxter robot as more positive and acceptable than NAO (Cuan et al., 2018). NAO was reportedly used along with Aibo in gesture experiments (Andry et al., 2011), iCub in eliciting behaviors on humans (Anzalone et al., 2015), Wifibot to carry the NAO (Canal et al., 2016), Pepper in human head imitation (Cazzato et al., 2019), Turtelbot in providing elderly care (DiMaria et al., 2017), Robokind R25 in interviewing humans (Henkel et al., 2019), Reeti (Johal et al., 2014) in expressing different parenting styles, R3 (Kose et al., 2014) in performing sign language gestures, Palro and Gemini (Pan et al., 2013) in evaluating interaction styles, and PR2 in identifying preferred human-robot proxemics (Rajamohan et al., 2019).

5.7 Novel Developments in Human-NAO Interaction

NAO has been used for unique purposes, which paved the way for new developments in human-robot interaction. One of the limitations of NAO is linked to physical abilities. Therefore, researchers try to improve physical contact with humans based on sensory information coming from them. Technical studies demonstrate promising results in relation to kinesthetic teaching by humans (Cho and Jo, 2011; Stadler et al., 2014). For instance, Bellaccini et al. (2014) proposed manual guidance of NAO without force sensors to improve physical human-robot interaction (pHRI). In a quite similar way, Berger et al. (2013) introduced a machine learning approach that enables NAO to follow human guidance by identifying human forces during a joint transportation task. Cao et al. (2014, 2017) presented a novel collaborative behavior controller ROBEE that selects actions based on homeostatic drive theory for NAO to jointly perform a task with participants more autonomously. In other words, this controller allows NAO to be aware of users' psychological (e.g., valence) and physical (e.g., thirst) needs. The brain-machine interface (BMI or BCI) has been commonplace in studies that address the problems of people with motor disabilities. Accordingly, some researchers proposed a novel BMI interface such as EOG/ERP hybrid human-machine interface (Ma et al., 2013), EEG-based recognition of imaginary movements of fingers

(Stankevich and Sonkin, 2016) and Emotive EPOC (Gomilko et al., 2016) to control NAO behaviours through commands by translating human brain activity. These findings show that NAO's limitations might be overcome by using more advanced deep learning solutions that enable the robot to function in natural environments.

5.8 From Close-Loop Systems Towards Real Autonomy

The realization of versatile and human-like intelligence and cognitive modules in robots remains a challenge for the HRI field. As shown by our analysis, all autonomous systems used in user studies were targeting a specific application field. Generic reasoning relates to developmental robotics that can include various theories and methods such as deep learning, sensorimotor information processing, metacognition, memory, and decision making (Miyazawa et al., 2019). Several studies proposed a cognitive architecture for NAO's system design. For instance, Adam et al. (2016) presented Cognitive and Affective Interaction-Oriented architecture (CAIO) that allows NAO to perceive its environment multi-modally, to manipulate mental states, to respond emotionally, and to execute physical and verbal actions. Aly and Dugan (2018) proposed Experiential Robot Learning in which NAO must autonomously learn and gradually acquire knowledge and skills through experience in the real world, achieved through reinforcement learning. Dindo and Zambuto (2010) focused on a multi-instance learning algorithm for NAO to learn the word-to-meaning associations through visual perceptions. Andry et al. (2011) presented an artificial neural network control architecture that allows rhythm detection to build an internal reward for learning inspired by human behavior. It has implications on the quality of the interaction in which NAO is capable of predicting and following human actions. To endow NAO with more adaptive behaviours, Bertacchini et al. (2017) designed a cognitive architecture that consists of human identification, emotions and gestures recognition and exhibition, and speech sentiment analysis in customer-robot interaction. Using computational modeling, Cantucci and Falcone (2019) endowed NAO with social autonomy in which it serves the role of infoPoint assistant that helps users to find out the location of their point of interest (e.g., a restaurant) and how to get to the place. Quite similarly, through the Internet of Things framework (IoT), Mondal and Nandi (2017) created a customizable assistant by enabling NAO to perform daily tasks that its owner requests. To increase the emotional aspect of interaction, Chella et al. (2013) built the cognitive architecture of NAO based on perceptual, emotional, and behavioural data. Another attempt in this area is made by Ribeiro et al. (2016) that presented the Socially Expressive Robotics Architecture (SERA) ecosystem for NAO as an autonomous and emphatic robot tutor in teaching sustainable development. These multiple examples of cognitive architectures for NAO are important to enable human-like intelligence and develop more natural HRI. A more detailed overview of research on cognitive architectures can be found in Ye et al. (2018) and Kotseruba et al. (2020).

5.9 Concluding Remarks

NAO is a well-accepted social robot valued for its fun and enjoyable appearance. However, there were mixed feelings about its interaction capabilities, which manifest diverse individual preferences and perceptions. Its interactive abilities can be empowered when displaying and recognizing emotions. Commonly, its body language is a medium for expressing emotions. NAO can detect emotions from facial expressions, and therefore, there is an emotional contagion in which NAO adapts to a human's emotional state or vice versa (Xu et al., 2015; Stojanovska et al., 2018). Users also want NAO to feel and show different kinds of emotions. For instance, students thought they wanted NAO to "feel life" and feel happiness and togetherness when interacting with them (Omokawa and Matsuura, 2018). As compared to the unemotional NAO, the emotional robot was considered more anthropomorphic, while its intelligence may not affect the perception of anthropomorphism (Zlotowski et al., 2014).

NAO is widely accepted as a socially assistive robot, which communicates with users socially rather than physically (Sarabia et al., 2018). A great body of research has used NAO as a mediator in autism therapy and other therapeutic interventions with older people. Using social robots can offer alternative or complementary ways to support traditional treatment. As a viable approach to autism treatment, robot-mediated autism intervention is designed to improve children's verbal and non-verbal behaviours as well as social skills. Levels of autism are known to be the most defining factor that accounts for different social interaction experiences and engagement rates (Ahmad et al., 2017). So far, the autism studies with NAO found that it has a great potential in helping children with autism to maintain eye contact (Anzalone et al., 2015), prefer specific instructions over spontaneity (Arent and Kruk-Lasocka, 2019) and augment communication skills (Hamid et al., 2013). Some other therapies focus on physical therapy, for instance, to improve motor learning skills of children with cerebral palsy (Rahman et al., 2015; Buitrago et al., 2020). Children with motor disabilities may become motivated and encouraged to do imitation and motor learning tasks. In addition, hearing-impacted children's sound detection improved over sessions, meaning that NAO can be used for auditory-verbal therapy (Ioannou and Andreva, 2019). Verbal communication with NAO has occurred in different learning and communication scenarios. Its speech is mainly based on scripted texts and therefore usually lacks personalized responses. Thus, autonomous and natural human-NAO verbal interaction is still at its infancy.

Users liked the robot's nonjudgemental behavior (Da Silva et al., 2018), and they were more engaged when the robot asked for personal details than quiz-like questions (Eyssel et al., 2017). In particular, game-based relationship with the robot may result in more self-disclosure (Groot et al., 2019). Furthermore, NAO was seen as more trustworthy and persuasive when compared to a virtual agent in either voice or virtual embodiment (Artstein et al., 2017). This distinctive characteristic hints that NAO can deal with sensitive issues carefully without making people feel uncomfortable when revealing oneself.

It was found that robots playing games with humans have an entertainment value (Johnson et al., 2016). Especially, it holds true for young children since their learning is mainly based on free play activities than instructed guidance on task performance. For instance, users preferred the R3 robot for learning, while NAO was associated with play (Kose et al., 2014). In another study, NAO played a board game known as tic-tac-toe with users on a tablet and showed that its behaviors could be robust with the help of voice synthesis and recognition. A more active, interactive, and extrovert robot is preferred as a partner in meetings (Mubin et al., 2014). There was no significant difference in user enjoyment between the system conditions, but most children tend to favor autonomous robot (Tozadore et al., 2017). Learning with NAO is interesting for children, and the content of play may affect the result of the learning process. Character recognition also plays an important role, how NAO recognises the kids' writing, and it can be spelled back towards them. In this case, the kids can learn how to pronounce English much better and learn the handwriting of the alphabet (Kim et al., 2019). The two-way communication is found to be effective since each child can respond to the questions from NAO (Miskam et al., 2013).

Personalization is a much-needed feature for all social robots, and NAO's case is no exception. It is commonly regarded that the robot may become more effective and human-like when it is able to tailor to user's needs and preferences and build a sustainable and long-term relationship. Personalized human-robot interactions are specifically suitable when robots interact with humans for longer periods (Irfan et al., 2019). In such context, robots may develop a certain kind of memory storage that allows them to remember and record all available information about people through continuous interaction with humans. Considering the variation in autism symptoms, there is a clear need for robot personalization in autism therapy (Fuglerud and Solheim, 2018). To illustrate, Greczek et al. (2014) emphasized that varied feedback may be more effective and less discouraging than descriptive feedback in an imitation game for children with autism. Also, Mirnig et al. (2011) found that human-robot interaction might be affected due to the provision or withholding of feedback. Users' perception of the robots could be distinguished based on different interaction styles even when it is a short-lived encounter (Pan et al., 2013). We come back to this subject later in the paper.

Personal characteristics of NAO are essential as each human shows idiosyncratic preferences in behaviours. What is interesting is that both introverted and extroverted humans wanted to interact with the personality-matching robot (Aly and Tapus, 2013). This posits that the personality traits of the robot are a relatively significant factor in relation to its non-verbal behavior. Users prefer to have a robot partner that shares the same personality as in the human-human interaction. Not surprisingly, it is suggested that extroverted robots positively affect interaction flow (Celiktutan and Gunes, 2015).

Similar to a human-human relationship, it may not be realistic if the human-robot interaction imitates faultless and impeccable communication. In psychology, *the pratfall effect* explains that a mistake would increase the interpersonal appeal and make humans more likable (Aronson et al., 2014). In this regard, Mirnig et al. (2017) highlights that the same phenomenon can be applied to social robots. In their study, participants liked the faulty robot significantly better than the robot that interacted flawlessly. The timing of the errors might also play an important role. Much interestingly, Lucas et al. (2018) found that NAO having conversational errors during warm-up conversation may recover sooner. Nevertheless, some users may develop biases toward the robot to be faulty and have limited skills (Turp et al., 2019). Although an erroneous robot is generally understudied, it certainly is one of the key areas to understand human-robot interaction in an unrestricted way.

Researchers have used external measurement devices such as RGB-D camera, eye tracker, motion detector, and many other tools for some decades. They make it possible to measure human features such as body postures, movement, speech, and gaze in a more accurate and reliable way. They can fill the gap in the robot's capabilities in measuring a human's input and engagement. In Craig et al. (2016), gaze tracking hardware is used to create gaze-based language command in order to facilitate the communication barriers between NAO and users. In another study, a speech recognition system called Cloud technology was used to assess the correct understanding of Chinese language words that were transmitted to NAO (Han et al., 2018). Other researchers use gesture recognition system based on external cameras (Ajili et al., 2017) and object detection algorithm to recognize the face from NAO's main camera (Cheng et al., 2012; Cuijpers and van der Pol, 2013). These advancements are significant as service robots are becoming popular in our domestic and social lives. In some way, it would be innovative if these technologies could also evaluate the quality of human-robot interaction. For instance, there might be some level of subjectivity in coding behaviours, especially in autism therapy (Baraka et al., 2017).

Existing research studies found no conclusive evidence regarding the benefits of social robots over other technologies. NAO's advantage over other technologies is still unclear as there are insufficient evidence for its benefit compared to tablets and computers. It might be intuitive to consider that users prefer to interact with a physical robot because of its animate and lively appearance. However, a user preference may depend on other factors, such as age and context. Notably, older adults who have serious underlying health issues may be inclined towards physical robots. For example, elderly people preferred robots over a tablet, despite technical limitations of the NAO (Olde Keizer et al., 2019). Furthermore, students perceived NAO as a sociable agent and favored it over other learning aids, e.g., a computer (Liles and Beer, 2015). Focusing on a language learning context, Zhexenova et al. (2020) reported that there is no significant difference in children's perceptions of NAO's effectiveness in comparison with a tablet and a human teacher. In the entertainment area, Wong et al. (2018) revealed that physically present NAO improved information dissemination and hence increased visibility of the advertised product.

6 KEY INSIGHTS: STRENGTHS AND LIMITATIONS

Our overall impression of the current review demands a further reflection on how research could be conducted with a social robot NAO. Although some points may be generic, we believe the research-based insights will benefit researchers either working or intending to work with NAO.

6.1 Strengths

NAO is commonly regarded as a widely used platform. Researchers tend to skip the details of why they choose NAO over other social robots except acknowledging its wider use. The most straightforward reason is that NAO has become the standard platform for RoboCup, an international robotics competition, for over 10 years.

NAO enjoys a great appeal from its end-users. Its child-like and non-threatening appearance makes it appealing. In particular, children at younger ages appear to engage with NAO more successfully than those at later childhood stages. This idea is supported by the significant number of studies that have been conducted in the educational and therapeutic context.

NAO is certainly not just an eye-catcher robot as its portability is highly appreciated by the researchers. Its small size in addition to light weight is helpful for easy transportation in a standard car (e.g. a taxi) which makes *in the wild* research possible.

NAO can be regarded as a plug-and-play robot due to its robust and easy setup characteristics. This allows researchers to have a reliable platform for a real-world deployment as confirmed by numerous research works conducted in diverse settings, ranging from schools to hospitals.

NAO is an affordable robot with a cost of around 6000 Euro[11]. Although it might be more expensive in comparison to other smaller humanoid robots, NAO is one of the most complete humanoid robots on the market in terms of functional and technical abilities.

NAO's customizable features also meet the needs of multi-disciplinary researchers worldwide. This is surely thanks to the multi-level programming framework proposed to researchers. While the block-based programming framework, Choregraphe, allows researchers from social sciences to easily implement novel behaviors, the C++/Python API allows engineers to develop novel technical contributions (i.e. computer vision, control, etc.) and deploy directly on the robot. The HRI field being so multi-disciplinary, its programming framework positively contributed to the success of the NAO platform.

NAO is multimodal in both its input and output communication modalities. It is relatively well equipped with internal sensors to perceive its environment as well as external actuators to perform verbal and non-verbal behaviors (e.g. body motion and LEDs).

NAO can take on a unique social role of one's learner. NAO as an educational robot has assisted poorly performing children to engage in a learning activity by taking up a unique social role of their learner. This can positively affect meta-cognitive abilities such as increased self-confidence and problem-solving (Hood et al., 2015). Other notable examples include handwriting practicing, second language learning, and studying school subjects like mathematics and science classes. With remarkable achievements in education, NAO is not much used in traditional and formal learning settings and rather acts as a one-to-one tutor, peer, or a learner (Johal, 2020).

NAO can bring cognitive and affective values when interacting with humans that have social and learning barriers. Although the robot can not replace the key social actors such as therapists and teachers, it can make learning and therapy engaging and fun experience, while educators can focus on creative as well as differentiated teaching practices.

NAO could be a great help for individuals who have less social experience and companionship in their life. For instance, in treating dementia and for other elderly care therapies, it could be applied to assist in daily life by monitoring and reminding to take the pills and do physical exercises following a certain plan instructed by medical staff. NAO as a companion may enhance the quality of life that most people expect to enjoy in their later lives.

Gendered stereotypes seem to persist in human-robot interaction. Multiple research indicate that users may perceive the robot in different ways based on gender markers such as voice and verbal commands (Sandygulova and O'Hare, 2015; Jackson et al., 2020). To a great extent, NAO is among the genderless robots (Obaid et al., 2016) compared to other robots (e.g., Kaspar, Milo). Thus, research with less gendered robots is important to eliminate gendered attitudes towards feminine and masculine qualities, which appear to contribute to the interaction outcomes.

6.2 Weaknesses

NAO has a low battery life and overheating issues that make it less durable than other social robots (e.g., Pepper). Generally, it works for 60 min in active use and 90 min in normal use. These issues question its sustainability and long-term efficacy. As our review shows, the majority of experiments with NAO usually happen on a short-term basis lasting for no more than 30 min. For instance, some participants are concerned with the robot being not active and responsive as they expected it to be. With that in mind, the activities and experimental design need to be adjusted time-wise.

Although NAO is relatively well equipped to perform near-human actions, it is quite often supported by input/output external devices such as high-quality or depth cameras and microphones. While NAO has two internal cameras, the low resolution does not allow to perform complex vision recognition tasks. For example, the closer a person is, the better the robot detects facial expressions and other visual cues, while it cannot recognize people who are more than 2 m away (Bolotnikova et al., 2017). Oftentimes, the use

[11]https://www.generationrobots.com/en/403100-programmable-humanoid-robot-nao-v6.html

of additional equipment such as touchscreens, tablets, or wearables can substitute for perceptual limitations (Johal, 2020).

NAO can hardly function in a noisy environment and recognize human speech. User's age influences speech recognition as young children and older people have different speech characteristics and coherence (Vargas et al., 2021). In particular, it is not yet practicable for NAO to recognize children's speech (Kennedy et al., 2017). Alternatively, researchers could use Google Cloud Speech recognition services that allow NAO understand different languages and optimize its workflow.

Hard surfaces are needed for NAO's movements and stable positioning. Aldebran first designed NAO as bipedal robot to walk in open loop. Closed loop walk algorithm was adopted on NAO humanoids that became capable of omnidirectional walking (Gouaillier et al., 2010; Kasari et al., 2019). NAO has a particular way of walking, and while the robot can move freely on flat and hard surfaces, it lacks robustness on surfaces such as on carpets or rugs (Shamsuddin et al., 2011). For instance, RoboCup teams like Berlin United (previously NAO Team Humboldt) have long been exploring the robot's ability to move and kick the soccer ball autonomously based on visual spatial perception[12].

Autonomy is undoubtedly the most demanding feature that most social robots lack. NAO has been predominantly used in the Wizard of Oz approach, a frequently employed method; wherein the interaction is controlled remotely by human input along the autonomy spectrum (Riek, 2012). Scripted, although highly constrained interactions are also commonly used solutions.

7 FUTURE RESEARCH WITH NAO

Our results allow us to make a number of recommendations for future research using NAO:

Data-driven behavior generation: While rule-based behaviour generation approaches perform well, they are often costly, time-consuming and bound up to expert knowledge. The cost of creating production rules and the need for manual configurations in order to generate complex and natural human behaviours put a limit to the complexity and diversity of generated behaviours. Thus, the development of data-driven behaviour generating systems using machine learning have to become the research focus as the actual human-human interaction data can provide a more human-like and multimodal behaviour generation (*see* Liu et al. (2021) for a review on gesture generation).

Long-term engagement: Although cross-sectional studies are commonplace due to different technological and methodological constraints, it is feasible to commit to long-term goals and test the efficacy of NAO and its capabilities. The user studies in robot-assisted educational and therapeutic settings need convincing evidence of the robot's long-term efficacy, especially those working with underserved populations (Rakhymbayeva et al., 2021).

Multi-party interaction: It would be suitable to observe and refine NAO's behaviors and its relationship with study participants in the presence of co-present others. One-on-one interaction has long been practiced, however, it is still unclear how NAO interacts with multiple participants. This type of interaction deserves much attention because it allows to maintain collaborative HRI. The robot's mediating role is important to facilitate human relationships such as student-student, student-tutor, and child-parent. In addition, professionals from other fields such as psychology and education can also contribute to evaluating the quality of human-robot interaction. For instance, in an educational setting, teachers may assess the interaction outcomes based on rubrics and observation.

Natural communication: Social dialogues should be more uplifting and engaging using more affective reactions. They may be based on a real interaction scenario where different participants react in varying ways. Interaction roles might be specified in advance, or users may find out in the course of the dialogue. Open-ended interactions can be designed where the robot is faulty or make errors during the interaction from which they can recover during the session. However, it might be helpful to maintain a cooperative imagined contact relying on real-life scenario. Research shows that imagined contact may provide humans with a behavioral guide, which probably improves their feelings of self-efficacy and confidence in future interaction (Kuchenbrandt and Eyssel, 2012).

Personalization: One cannot fit all, especially when it comes to social interaction. For that reason, it seems that adaptation and personalization have been under investigated as the NAO robot was used across various populations and cultures without much changes. Interventions have to be aware of user demographics which is the most straightforward way to adapt the content by adding specific verbal and non-verbal cues. The decision over how much personalization to use has to derive from study focus and population, which is highly anticipated of any experiment. In the case of young children with autism, there is a strong need for customized robot behaviors, as these children show varying degrees of autism symptoms that develop individually. For this reason, the NAO can target different social skills development activities and then find out what works best for a certain child. It would be an important objective for NAO to learn child preferences from session to session and adapt its behaviors accordingly.

Impact of COVID-19 on HRI: If we consider the significant decrease in an experiment-based HRI, it becomes clear that some of us may not embrace an online research environment. There might be a serious disparity between subject areas, institutional support, and early-career and expert researchers. Besides, there is a geographical factor that might influence research activity as some countries (e.g. Israel, New Zealand) cope better with COVID-19, while others (e.g. USA, Italy) have been hardest hit by it. Thus, a collaboration between researchers within and beyond the field can be a silver lining of current research-related challenges.

8 CONCLUSION AND LIMITATIONS

In HRI, we often work and develop closer ties with a particular robot, and may overlook how other robots contribute to the field. In this review, we presented a comprehensive overview on the use of NAO, which is a remarkable social robot in many instances. So far, NAO has been exposed to challenging yet rewarding journey.

[12]https://www.naoteamhumboldt.de/en/team/

Its social roles have expanded thanks to its likeable appearance and multi-modal capabilities followed by its fitness to deliver socially important tasks. Still, there are gaps to be filled in view of sustainable and user-focused human-NAO interaction. We hope that our review can contribute to the field of HRI that needs more reflection and general evidence on the use of the social robots, such as NAO in a wide variety of contexts. The main limitation to this study is that our search was limited to keywords in abstract and titles. It means that we could not cover other studies that might enrich the current review. Nevertheless, we believe that our research may engender important insights into the use of NAO across different domains and shape a broader understanding of human-robot interaction over the last decade. An implication of the findings shows a greater need for increasing the value and practical application of NAO in user-centered studies. Future studies should consider the importance of real-world and unrestricted experiments with NAO and involve other humans that might facilitate human-robot interaction.

DATA AVAILABILITY STATEMENT

The dataset generated for this study can be found in the Zenodo repository https://zenodo.org/record/5576799.

AUTHOR CONTRIBUTIONS

AA—annotation, writing; NR—annotation, visualization; EY - annotation, writing, review and editing; AS - conceptualization, supervision, funding acquisition, writing, review and editing; and WJ—conceptualization, supervision, funding acquisition, data visualisation, writing, review and editing.

FUNDING

This work was supported by the Nazarbayev University Collaborative Research Program grant (091019CRP2107), the Australian Research Council (DE210100858) and the GROW grant from the Faculty of Engineering UNSW.

ACKNOWLEDGMENTS

We would like to thank the authors for kindly sharing their articles in closed access.

REFERENCES

Adam, C., Johal, W., Pellier, D., Fiorino, H., and Pesty, S. (2016). "Social Human-Robot Interaction: A New Cognitive and Affective Interaction-Oriented Architecture," in *Social Robotics*. Editors A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He (Cham: Springer International Publishing), 253–263. doi:10.1007/978-3-319-47437-3_25

Ahmad, M. I., Shahid, S., and Tahir, A. (2017). "Towards the Applicability of NAO Robot for Children with Autism in Pakistan," in Human-Computer Interaction - INTERACT 2017 - 16th IFIP TC 13 International Conference of Lecture Notes in Computer Science, Mumbai, India, September 25–29, 2017. Editors R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler (Springer), 463–472. doi:10.1007/978-3-319-67687-6_32

Ajili, I., Mallem, M., and Didier, J.-Y. (2017). "Gesture Recognition for Humanoid Robot Teleoperation," in 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), Lisbon, Portugal, August 28–31, 2017, 1115–1120. doi:10.1109/roman.2017.8172443

Alenljung, B., Andreasson, R., Lowe, R., Billing, E., and Lindblom, J. (2018). Conveying Emotions by Touch to the Nao Robot: A User Experience Perspective. *Mti* 2, 82. doi:10.3390/mti2040082

Ali, S., Mehmood, F., Ayaz, Y., Asgher, U., and Khan, M. J. (2020). "Effect of Different Visual Stimuli on Joint Attention of Asd Children Using Nao Robot," in *Advances in Neuroergonomics and Cognitive Engineering*. Editor H. Ayaz (Cham: Springer International Publishing), 490–499. doi:10.1007/978-3-030-20473-0_48

Aly, A. A., and Dugan, J. B. (2018). "Experiential Robot Learning with Deep Neural Networks," in 7th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, Lisbon, Portugal, September 18–21, 2017 (New York City, NY: ICDL-EpiRob), 356–361. doi:10.1109/DEVLRN.2017.8329830

Aly, A., and Tapus, A. (2013). "A Model for Synthesizing a Combined Verbal and Nonverbal Behavior Based on Personality Traits in Human-Robot Interaction," in ACM/IEEE International Conference on Human-Robot Interaction, Tokyo, Japan, March 3–6, 2013, 325–332. doi:10.1109/hri.2013.6483606

Anastasiou, D., Jokinen, K., and Wilcock, G. (2013). "Evaluation of WikiTalk - User Studies of Human-Robot Interaction," in *Human-Computer Interaction. Interaction Modalities and Techniques*. Editor M. Kurosu (Berlin, Heidelberg: Springer Berlin Heidelberg), 32–42. doi:10.1007/978-3-642-39330-3_4

Andry, P., Blanchard, A., and Gaussier, P. (2011). Using the Rhythm of Nonverbal Human-Robot Interaction as a Signal for Learning. *IEEE Trans. Auton. Ment. Dev.* 3, 30–42. doi:10.1109/TAMD.2010.2097260

Ansermin, E., Mostafaoui, G., Sargentini, X., and Gaussier, P. (2017). "Unintentional Entrainment Effect in a Context of Human Robot Interaction: An Experimental Study," in RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication 2017, Lisbon, Portugal, August 28–September 1, 2017, 1108–1114. doi:10.1109/ROMAN.2017.8172442

Anzalone, S. M., Boucenna, S., Ivaldi, S., Chetouani, M., Anzalone, S. M., Boucenna, S., et al. (2015). Evaluating the Engagement with Social Robots. *Int. J. Soc. Robotics* 7, 465–478. doi:10.1007/s12369-015-0298-7

Arent, K., Kruk-Lasocka, J., Niemiec, T., and Szczepanowski, R. (2019). "Social Robot in Diagnosis of Autism Among Preschool Children," in 2019 24th International Conference on Methods and Models in Automation and Robotics, Miedzyzdroje, Poland, August 26–29, 2019 (New York City, NY: MMAR), 652–656. doi:10.1109/MMAR.2019.8864666

Arias-Aguilar, J. A., Palacios-Villavicencio, M. L., Bretado-Gallegos, R., Medina-Nieto, M. A., Ruiz, A. B., Rodríguez-López, V., et al. (2017). "Analysis of Children," in Proceedings of the XVIII International Conference on Human Computer Interaction - Interacción '17, New York, NY, Florida City, FL, May 22–24, 2017 (New York City, NY: ACM Press), 1–4. doi:10.1145/3123818.3123837

Aronson, E., Willerman, B., and Floyd, J. (2014). The Effect of a Pratfall on Increasing Interpersonal Attractiveness. *Psychon. Sci.* 4, 227–228. doi:10.3758/BF03342263

Artstein, R., Traum, D., Boberg, J., Gainer, A., Gratch, J., Johnson, E., et al. (2017). "Listen to My Body: Does Making Friends Help Influence People in FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference, 430–435.

Balkibekov, K., Meiirbekov, S., Tazhigaliyeva, N., and Sandygulova, A. (2016). "Should Robots Win or Lose? Robot's Losing Playing Strategy Positively Affects Child Learning," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication, New York City, NY, August 26–31, 2016 (RO-MAN (IEEE), 706–711.

Baraka, K., Alves-Oliveira, P., and Ribeiro, T. (2020). "An Extended Framework for Characterizing Social Robots," in *Human-Robot Interaction* (Springer), 21–64. doi:10.1007/978-3-030-42307-0_2

Baraka, K., Melo, F. S., and Veloso, M. (2017). "'Autistic Robots' for Embodied Emulation of Behaviors Typically Seen in Children with Different Autism Severities,". *Social Robotics*. Editors A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, et al. (Cham: Springer International Publishing), 105–114. doi:10.1007/978-3-319-70022-9_11

Bechade, L., Dubuisson Duplessis, G., Sehili, M., and Devillers, L. (2015). "Behavioral and Emotional Spoken Cues Related to Mental States in Human-Robot Social Interaction," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15, Seattle, WA, November 9–13, 2015 (New York City, NY: ACM Press), 347–350. doi:10.1145/2818346.2820777

Beck, A., Stevens, B., Bard, K. A., and Cañamero, L. (2012). Emotional Body Language Displayed by Artificial Agents. *ACM Trans. Interact. Intell. Syst.* 2, 1–29. doi:10.1145/2133366.2133368

Bellaccini, M., Lanari, L., Paolillo, A., and Vendittelli, M. (2014). Manual Guidance of Humanoid Robots without Force Sensors: Preliminary Experiments with NAO. *Proc. - IEEE Int. Conf. Robotics Automation* 8, 1184–1189. doi:10.1109/ICRA.2014.6907003

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social Robots for Education: A Review. *Sci. Robot* 3, eaat5954. doi:10.1126/scirobotics.aat595410.1126/scirobotics.aat5954

Berger, E., Vogt, D., Haji-Ghassemi, N., Jung, B., and Amor, H. (2013). "Inferring Guidance Information in Cooperative Human-Robot Tasks," in IEEE-RAS International Conference on Humanoid Robots, Atlanta, GA, October 15–17, 2013, 124–129. doi:10.1109/HUMANOIDS.2013.7029966

Bertacchini, F., Bilotta, E., and Pantano, P. (2017). Shopping with a Robotic Companion. *Comput. Hum. Behav.* 77, 382–395. doi:10.1016/j.chb.2017.02.064

Bethel, C. L., Eakin, D., Anreddy, S., Stuart, J. K., and Carruth, D. (2013). "Eyewitnesses Are Misled by Human but Not Robot Interviewers," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Tokyo, Japan, March 3–6, 2013 (IEEE), 25–32. doi:10.1109/HRI.2013.6483498

Bethel, C. L., and Murphy, R. R. (2010). Review of Human Studies Methods in Hri and Recommendations. *Int. J. Soc. Robotics* 2, 347–359. doi:10.1007/s12369-010-0064-9

Bevan, C., and Stanton Fraser, D. (2015). "Shaking Hands and Cooperation in Tele-Present Human-Robot Negotiation," in ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, March 2–5, 2015, 247–254. doi:10.1145/2696454.2696490

Bolotnikova, A., Demirel, H., and Anbarjafari, G. (2017). Real-time Ensemble Based Face Recognition System for NAO Humanoids Using Local Binary Pattern. *Analog Integr. Circ. Sig Process.* 92, 467–475. doi:10.1007/s10470-017-1006-3

Buitrago, J. A., Bolaños, A. M., and Caicedo Bravo, E. (2020). A Motor Learning Therapeutic Intervention for a Child with Cerebral Palsy Through a Social Assistive Robot. *Disabil. Rehabil. Assistive Technology* 15, 357–362. doi:10.1080/17483107.2019.1578999

Canal, G., Escalera, S., and Angulo, C. (2016). A Real-Time Human-Robot Interaction System Based on Gestures for Assistive Scenarios. *Computer Vis. Image Understanding* 149, 65–77. doi:10.1016/j.cviu.2016.03.004

Cantucci, F., and Falcone, R. (2019). A Computational Model for Cognitive Human-Robot Interaction: An Approach Based on Theory of Delegation. *CEUR Workshop Proc.* 2404, 127–133.

Cao, H.-L., Esteban, P., De Beir, A., Simut, R., Van De Perre, G., Lefeber, D., et al. (2014). "ROBEE: A Homeostatic-Based Social Behavior Controller for Robots in Human-Robot Interaction Experiments," in IEEE International Conference on Robotics and Biomimetics, Bali, Indonesia, December 5–10, 2014, 516–521. doi:10.1109/ROBIO.2014.7090383

Cao, H.-L., Gómez Esteban, P., Albert, D. B., Simut, R., Van de Perre, G., Lefeber, D., et al. (2017). A Collaborative Homeostatic-Based Behavior Controller for Social Robots in Human-Robot Interaction Experiments. *Int. J. Soc. Robotics* 9, 675–690. doi:10.1007/s12369-017-0405-z

Cazzato, D., Cimarelli, C., Sanchez-Lopez, J. L., Olivares-Mendez, M. A., and Voos, H. (2019). Real-time Human Head Imitation for Humanoid Robots. *ACM Int. Conf. Proceeding Ser.*, Singapore, July 27–29, 2019 65–69. doi:10.1145/3348488.3348501

Celiktutan, O., and Gunes, H. (2015). "Computational Analysis of Human-Robot Interactions Through First-Person Vision: Personality and Interaction Experience," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Kobe, Japan, August 31–September 4, 2015, 815–820. doi:10.1109/ROMAN.2015.7333602

Chella, A., Sorbello, R., Pilato, G., Vassallo, G., and Giardina, M. (2013). "A New Humanoid Architecture for Social Interaction between Human and a Robot Expressing Human-like Emotions Using an Android mobile Device as Interface," in *Biologically Inspired Cognitive Architectures 2012*. Editors A. Chella, R. Pirrone, R. Sorbello, and K. R. Jóhannsdóttir (Berlin, Heidelberg: Springer Berlin Heidelberg), 95–103. doi:10.1007/978-3-642-34274-5_21

Chevalier, P., Raiola, G., Martin, J.-C., Isableu, B., Bazile, C., and Tapus, A. (2017). "Do Sensory Preferences of Children with Autism Impact an Imitation Task with a Robot in ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, March 6–9, 2017, 177–186. doi:10.1145/2909824.3020234

Cho, S., and Jo, S. (2011). "Kinesthetic Learning of Behaviors in a Humanoid Robot," in International Conference on Control, Automation and Systems, Gyeonggi-do, South Korea, October 26–29, 2011, 1108–1112.

Craig, T. L., Nelson, C. A., Li, S., and Zhang, X. (2016). "Human Gaze Commands Classification: A Shape Based Approach to Interfacing with Robots," in 12th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), Auckland, New Zealand, August 29–31, 2016. doi:10.1109/MESA.2016.7587154

Cruz-Maya, A., Ferland, F., and Tapus, A. (2015). Social Facilitation in a Game-like Human-Robot Interaction Using Synthesized Emotions and Episodic Memory. *ICSR*, 164–173. doi:10.1007/978-3-319-25554-5_17

Cuan, C., Pakrasi, I., Berl, E., and LaViers, A. (2018). "CURTAIN and Time to Compile: A Demonstration of an Experimental Testbed for Human-Robot Interaction," in 2018 27th IEEE International Symposium on Robot and Human Interactive Communication Nanjing, China, August 27–31, 2018 (RO-MAN IEEE), 255–261. doi:10.1109/roman.2018.8525520

Cuijpers, R. H., and van der Pol, D. (2013). "Region of Eye Contact of Humanoid Nao Robot Is Similar to that of a Human," in *Social Robotics*. Editors G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards (Cham: Springer International Publishing), 280–289. doi:10.1007/978-3-319-02675-6_28

Das, D., Rashed, M. G., Kobayashi, Y., and Kuno, Y. (2015). Supporting Human-Robot Interaction Based on the Level of Visual Focus of Attention. *IEEE Trans. Human-mach. Syst.* 45, 664–675. doi:10.1109/THMS.2015.2445856

De Groot, J.-J., Barakova, E., Lourens, T., van Wingerden, E., and Sterkenburg, P. (2019). "Game-based Human-Robot Interaction Promotes Self-Disclosure in People with Visual Impairments and Intellectual Disabilities," in Understanding the Brain Function and Emotions - 8th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC Almeria, Spain, June 3–7, 2019, 262–272. doi:10.1007/978-3-030-19591-5_27

de Wit, J., Brandse, A., Krahmer, E., and Vogt, P. (2020). "Varied Human-like Gestures for Social Robots," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, March 23–26, 2020 (New York City, NY: ACM), 359–367. doi:10.1145/3319502.3374815

De Beir, A., Cao, H.-L., Gómez Esteban, P., Van de Perre, G., Lefeber, D., and Vanderborght, B. (2016). Enhancing Emotional Facial Expressiveness on NAO. *Int. J. Soc. Robotics* 8, 513–521. doi:10.1007/s12369-016-0363-x

DI Nuovo, A., Bamforth, J., Conti, D., Sage, K., Ibbotson, R., Clegg, J., et al. (2020). "An Explorative Study on Robotics for Supporting Children with Autism Spectrum Disorder During Clinical Procedures," in ACM/IEEE International Conference on Human-Robot Interaction Cambridge, UK, March 23–26, 2020, 189–191. doi:10.1145/3371382.3378277

Diaz, S. S., Shaik, J. M. M., Santofimio, J. C. G., Quintero M., C. G., and Quintero, M., C. G. (2018). "Intelligent Execution of Behaviors in a NAO Robot Exposed to Audiovisual Stimulus," in IEEE 2nd Colombian Conference on Robotics and Automation (CCRA) (IEEE), Barranquilla, Colombia, November 1–3, 2018, 1–6. doi:10.1109/CCRA.2018.8588149

DiMaria, N., Tan, K., Salgado, B., Su, W., and Vesonder, G. (2017). "Eldercare Robotics Revolution-Explaining Robotics for Eldercare," in 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON) (IEEE), New York City, NY, October 19–21, 2017, 431–436. doi:10.1109/UEMCON.2017.8249030

Dindo, H., and Zambuto, D. (2010). "A Probabilistic Approach to Learning a Visually Grounded Language Model through Human-Robot Interaction," in IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings, Taipei, Taiwan, October 18–22, 2010, 790–796. doi:10.1109/IROS.2010.5654440

Erden, M. S. (2013). Emotional Postures for the Humanoid-Robot Nao. *Int. J. Soc. Robotics* 5, 441–456. doi:10.1007/s12369-013-0200-4

Eyssel, F., Wullenkord, R., and Nitsch, V. (2017). "The Role of Self-Disclosure in Human-Robot Interaction," in RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication, Lisbon, Portugal, August 28–September 1, 2017, 922–927. doi:10.1109/ROMAN.2017.817241310.1109/roman.2017.8172413

Franco, G. A. M. (2015). "Evaluation of the Emotional Answer in HRI on a Game Situation," in Proceedings of the Latin American Conference on Human Computer Interaction - CLIHC '15, Córdoba, Argentina, November 18–21, 2015 (New York City, NY: ACM Press), 1–7. doi:10.1145/2824893.2824897

Fuglerud, K. S., and Solheim, I. (2018). The Use of Social Robots for Supporting Language Training of Children. *Stud. Health Technol. Inform.* 256, 401–408. doi:10.3233/978-1-61499-923-2-401

Galvão Gomes da Silva, J., Kavanagh, D. J., Belpaeme, T., Taylor, L., Beeson, K., and Andrade, J. (2018). Experiences of a Motivational Interview Delivered by a Robot: Qualitative Study. *J. Med. Internet Res.* 20, e116. doi:10.2196/jmir.7737

Gomilko, S., Zimina, A., and Shandarov, E. (2016). Attention Training Game with Aldebaran Robotics Nao and Brain-Computer Interface. *Interactive Collaborative Robotics*, 27–31. doi:10.1007/978-3-319-43955-6_4

Goodrich, M. A., and Schultz, A. C. (2007). Human-robot Interaction: A Survey. *FNT in Human-Computer Interaction* 1, 203–275. doi:10.1561/1100000005

Gouaillier, D., Collette, C., and Kilner, C. (2010). "Omni-directional Closed-Loop Walk for Nao," in 2010 10th IEEE-RAS International Conference on Humanoid Robots, 448–454. doi:10.1109/ICHR.2010.5686291

Graña, M., and Triguero, A. (2019). An Approach to Teach Nao Dialogue Skills. *IWINAC*, 301–308. doi:10.1007/978-3-030-19591-5_31

Greczek, J., Kaszubski, E., Atrash, A., and Mataric, M. (2014). "Graded Cueing Feedback in Robot-Mediated Imitation Practice for Children with Autism Spectrum Disorders," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Edinburgh, UK, August 25–29, 2014, 561–566. doi:10.1109/roman.2014.6926312

Gurpinar, C., Alasag, T., and Kose, H. (2012). "Humanoid Robot painter Assisted by a Human," in 2012 IEEE Workshop on Advanced Robotics and its Social Impacts, Munich, Germany, May 21–23, 2012 (ARSO IEEE), 79–82. doi:10.1109/ARSO.2012.6213403

Hamid, A. C., Miskam, M. A., Yussof, H., Shamsuddin, S., Hashim, H., and Ismail, L. (2013). Human-robot Interaction (Hri) for Children with Autism to Augment Communication Skills. *Amm* 393, 598–603. doi:10.4028/www.scientific.net/amm.393.598

Han, Y., Zhang, M., Li, Q., and Liu, S. (2018). "Chinese Speech Recognition and Task Analysis of Aldebaran Nao Robot," in 2018 Chinese Control And Decision Conference, Shenyang, China, June 9–11, 2018 (CCDCIEEE), 4138–4142. doi:10.1109/CCDC.2018.8407843

Henkel, Z., Baugus, K., Bethel, C. L., and May, D. C. (2019). User Expectations of Privacy in Robot Assisted Therapy. *Paladyn* 10, 140–159. doi:10.1515/pjbr-2019-0010

Henschel, A., Laban, G., and Cross, E. S. (2021). What Makes a Robot Social? A Review of Social Robots from Science Fiction to a home or Hospital Near You. *Curr. Robot Rep.* 2, 9–19. doi:10.1007/s43154-020-00035-0

Hernandez-Cedeño, J., Ramírez-Benavides, K., Guerrero, L., and Vega, A. (2019). "NAO as a Copresenter in a Robotics Workshop - Participant's Feedback on the Engagement Level Achieved with a Robot in the Classroom," in *Advances in Human Factors in Robots and Unmanned Systems*. Editor J. Chen (Cham: Springer International Publishing), 143–152. doi:10.1007/978-3-319-94346-6_14

Hirose, J., Hirokawa, M., and Suzuki, K. (2015). "Gaming Humanoid: A Humanoid Video Game Player with Emotional/encouraging Movement and Skill Level Control," in 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), Kobe, Japan, August 31–September 4, 2015, 54–59. doi:10.1109/roman.2015.7333625

Hong Hu, H., Zhao, J., Hongbo Li, H., Li, W., and Chen, G. (2016). "Telepresence Control of Humanoid Robot via High-Frequency Phase-Tagged SSVEP Stimuli," in 2016 IEEE 14th International Workshop on Advanced Motion Control, Auckland, New Zealand, April 22–24, 2016 (New York City, NY: AMC), 214–219. doi:10.1109/AMC.2016.7496353

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "When Children Teach a Robot to Write," in ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, March 2–5, 2015, 83–90. doi:10.1145/2696454.2696479

Horstmann, A. C., and Krämer, N. C. (2020). Expectations vs. Actual Behavior of a Social Robot: An Experimental Investigation of the Effects of a Social Robot's Interaction Skill Level and its Expected Future Role on People's Evaluations. *PLoS ONE* 15, e0238133. doi:10.1371/journal.pone.0238133

Huskens, B., Palmen, A., Van der Werff, M., Lourens, T., and Barakova, E. (2015). Improving Collaborative Play Between Children with Autism Spectrum Disorders and Their Siblings: The Effectiveness of a Robot-Mediated Intervention Based on Lego Therapy. *J. Autism Dev. Disord.*, 45, 3746–3755. doi:10.1007/s10803-014-2326-0

Imbernón Cuadrado, L.-E., Manjarrés Riesco, Á., and De La Paz López, F. (2016). Artie: An Integrated Environment for the Development of Affective Robot Tutors. *Front. Comput. Neurosci.* 10. doi:10.3389/fncom.2016.00077

Ioannou, A., and Andreva, A. (2019). "Play and Learn with an Intelligent Robot: Enhancing the Therapy of Hearing-Impaired Children," in *Human-Computer Interaction – INTERACT 2019*. Editors D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris (Cham: Springer International Publishing), 436–452. doi:10.1007/978-3-030-29384-0_27

Irfan, B., Ramachandran, A., Spaulding, S., Glas, D. F., Leite, I., and Koay, K. L. (2019). "Personalization in Long-Term Human-Robot Interaction," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction, Daegu, South Korea, March 11–14, 2019 (New York City, NY: HRI), 685–686. doi:10.1109/HRI.2019.8673076

Ismail, L. I., Shamsudin, S., Yussof, H., Hanapiah, F. A., and Zahari, N. I. (2012). Robot-based Intervention Program for Autistic Children with Humanoid Robot NAO: Initial Response in Stereotyped Behavior. *Proced. Eng.* 41, 1441–1447. doi:10.1016/j.proeng.2012.07.333

Jackson, R. B., Williams, T., and Smith, N. (2020). "Exploring the Role of Gender in Perceptions of Robotic Noncompliance," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, March 23–26, 2020 (New York City, NY: Association for Computing Machinery HRI '20), 559–567. doi:10.1145/3319502.3374831

Jessup, S., Schneider, T., Alarcon, G., Ryan, T., and Capiola, A. (2019). *The Measurement of the Propensity to Trust Automation*. Springer International Publishing, 476–489. doi:10.1007/978-3-030-21565-1_32

Johal, W., Calvary, G., and Pesty, S. (2014). "A Robot with Style, Because You Are worth it!," in Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14, Toronto, ON, April 26–May 1, 2014 (New York City, NY: ACM Press), 1201–1206. doi:10.1145/2559206.2581229

Johal, W. (2020). Research Trends in Social Robots for Learning. *Curr. Robot Rep.* 1, 75–83. doi:10.1007/s43154-020-00008-3

Johnson, D. O., Cuijpers, R. H., Pollmann, K., and van de Ven, A. A. J. (2016). Exploring the Entertainment Value of Playing Games with a Humanoid Robot. *Int. J. Soc. Robotics* 8, 247–269. doi:10.1007/s12369-015-0331-x

Kaptein, F., Broekens, J., Hindriks, K., and Neerincx, M. (2017). "Personalised Self-Explanation by Robots: The Role of Goals versus Beliefs in Robot-Action Explanation for Children and Adults," in RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication, Lisbon, Portugal, August 28–September 1, 2017, 676–682. doi:10.1109/ROMAN.2017.8172376

Kasari, M. M., Lau, N., and Pereira, A. (2019). A Fast and Stable Omnidirectional Walking Engine for the Nao Humanoid Robot. *CoRR* abs/1906, 06932.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child Speech Recognition in Human-Robot Interaction," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, March 6–9, 2017 (New York City, NY: Association for Computing Machinery), 82–90. doi:10.1145/2909824.3020229

Kim, A., Omarova, M., Zhaksylyk, A., Asselborn, T., Johal, W., Dillenbourg, P., et al. (2019). "CoWriting Kazakh: Transitioning to a New Latin Script Using Social Robots," in 2019 28th IEEE International Conference on Robot and

Human Interactive Communication, New Delhi, India, October 14–18, 2019 (RO-MAN IEEE), 1–6. doi:10.1109/RO-MAN46459.2019.8956471

Kose, H., Akalin, N., and Uluer, P. (2014). Socially Interactive Robotic Platforms as Sign Language Tutors. *Int. J. Hum. Robot.* 11, 1450003. doi:10.1142/S0219843614500030

Kose, H., Yorganci, R., Algan, E. H., and Syrdal, D. S. (2012). Evaluation of the Robot Assisted Sign Language Tutoring Using Video-Based Studies. *Int. J. Soc. Robotics* 4, 273–283. doi:10.1007/s12369-012-0142-2

Kotseruba, I., Tsotsos, J. K., and Tsotsos, J. (2020). 40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications. *Artif. Intell. Rev.* 53, 17–94. doi:10.1007/s10462-018-9646-y

Kraus, J. M., Nothdurft, F., Hock, P., Scholz, D., Minker, W., and Baumann, M. (2016). "Human after All," in Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - Automotive'UI 16, Ann Arbor, MI, October 24–26, 2016 (New York City, NY: ACM Press), 129–134. doi:10.1145/3004323.3004338

Kraus, M., Kraus, J., Baumann, M., and Minker, W. (2018). "Effects of Gender Stereotypes on Trust and Likability in Spoken Human-Robot Interaction," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, May 7–12, 2018 (Paris, France: European Language Resources Association ELRA), 112–118.

Krogsager, A., Segato, N., and Rehm, M. (2014). "Backchannel Head Nods in Danish First Meeting Encounters with a Humanoid Robot: The Role of Physical Embodiment," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*. Editor M. Kurosu (Cham: Springer International Publishing), 651–662. doi:10.1007/978-3-319-07230-2_62

Kuchenbrandt, D., and Eyssel, F. (2012). "The Mental Simulation of a Human-Robot Interaction: Positive Effects on Attitudes and Anxiety toward Robots," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Paris, France, September 9–13, 2012, 463–468. doi:10.1109/ROMAN.2012.6343795

Liles, K. R., and Beer, J. M. (2015). "Rural Minority Students' Perceptions of Ms. an, the Robot Teaching Assistant, as a Social Teaching Tool," in Proceedings of the Human Factors and Ergonomics Society, Los Angeles, CA, October 26–30, 2015, 372–376. doi:10.1177/1541931215591077

Liu, Y., Mohammadi, G., Song, Y., and Johal, W. (2021). "Speech-based Gesture Generation for Robots and Embodied Agents: A Scoping Review," in 2021 9th International Conference on Human-Agent Interaction, Online, November 9–11, 2021 (New York City, NY: HAI '21 ACM). doi:10.1145/3472307.3484167

Liying Cheng, L., Qi Sun, Q., Han Su, H., Yang Cong, Y., and Shuying Zhao, S. (2012). "Design and Implementation of Human-Robot Interactive Demonstration System Based on Kinect," in Proceedings of the 2012 24th Chinese Control and Decision Conference, CCDC, Taiyuan, China, May 23–25, 2012, 971–975. doi:10.1109/CCDC.2012.6242992

Lopez-Rincon, A. (2019). "Emotion Recognition Using Facial Expressions in Children Using the NAO Robot," in 2019 International Conference on Electronics, Communications and Computers, Cholula, Mexico, Feburary 27–March 1, 2019 (CONIELECOMP IEEE), 146–153. doi:10.1109/CONIELECOMP.2019.8673111

Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A., et al. (2018). "Getting to Know Each Other," in ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, March 5–8, 2018, 344–351. doi:10.1145/3171221.3171258

Ma, J., Zhang, Y., Nam, Y., Cichocki, A., and Matsuno, F. (2013). "EOG/ERP Hybrid Human-Machine Interface for Robot Control," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, November 3–7, 2013, 859–864. doi:10.1109/iros.2013.6696451

Manohar, V., and Crandall, J. (2014). Programming Robots to Express Emotions: Interaction Paradigms, Communication Modalities, and Context. *IEEE Trans. Human-mach. Syst.* 44, 362–373. doi:10.1109/THMS.2014.2309662

Manome, N., Shinohara, S., Suzuki, K., and Mitsuyoshi, S. (2020). "Machine Translation from Japanese to Robot Language for Human-Friendly Communication," in *HCI International 2020 - Posters*. Editors C. Stephanidis and M. Antona (Cham: Springer International Publishing), 254–260. doi:10.1007/978-3-030-50726-8_33

Marin Vargas, A., Cominelli, L., Dell'Orletta, F., and Scilingo, E. P. (2021). Verbal Communication in Robotics: A Study on Salient Terms, Research fields and Trends in the Last Decades Based on a Computational Linguistic Analysis. *Front. Comput. Sci.* 2, 591164. doi:10.3389/fcomp.2020.591164

Mirnig, N., Riegler, S., Weiss, A., and Tscheligi, M. (2011). "A Case Study on the Effect of Feedback on Itinerary Requests in Human-Robot Interaction," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Atlanta, GA, July 31–August 3, 2011, 343–349. doi:10.1109/ROMAN.2011.6005260

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., and Tscheligi, M. (2017). To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Front. Robot. AI* 4, 21. doi:10.3389/frobt.2017.00021

Miskam, M. A., Hamid, M. A. C., Yussof, H., Shamsuddin, S., Malik, N. A., and Basir, S. N. (2013). Study on Social Interaction between Children with Autism and Humanoid Robot NAO. *Amm* 393, 573–578. doi:10.4028/www.scientific.net/amm.393.573

Miskam, M. A., Shamsuddin, S., Yussof, H., Ariffin, I. M., and Omar, A. R. (2015). "A Questionnaire-Based Survey: Therapist's Response on Emotions Gestures Using Humanoid Robot for Autism," in 2015 International Symposium on Micro-NanoMechatronics and Human Science, Nagoya, Japan, November 23–25, 2015 (New York City, NY: MHS), 1–7. doi:10.1109/mhs.2015.7438298

Miyazawa, K., Horii, T., Aoki, T., and Nagai, T. (2019). Integrated Cognitive Architecture for Robot Learning of Action and Language. *Front. Robot AI* 6, 131. doi:10.3389/frobt.2019.00131

Mondal, S., and Nandi, G. C. (2017). "Personal Robot: Towards Developing a Complete Humanoid Robot Assistant Using the Internet of Things," in 2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, San Francisco, CA, August 4–8, 2017. doi:10.1109/UIC-ATC.2017.8397453

Mubin, O., D'Arcy, T., Murtaza, G., Simoff, S., Stanton, C., and Stevens, C. (2014). "Active or Passive?: Investigating the Impact of Robot Role in Meetings," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Edinburgh, UK, August 25–29, 2014, 580–585. doi:10.1109/ROMAN.2014.6926315

Mubin, O., Stevens, C. J., Shahid, S., Mahmud, A. A., and Dong, J.-J. (2013). A Review of the Applicability of Robots in Education. *Technology Education Learn.* 1. doi:10.2316/Journal.209.2013.1.209-0015

Mwangi, E., Barakova, E., Zhang, R., Diaz, M., Catala, A., and Rauterberg, M. (2016). "See Where I Am Looking at," in HAI 2016 - Proceedings of the 4th International Conference on Human Agent Interaction, Biopolis, Singapore, October 4–7, 2016, 329–332. doi:10.1145/2974804.2980479

Natale, L., Paikan, A., Randazzo, M., and Domenichelli, D. E. (2016). The Icub Software Architecture: Evolution and Lessons Learned. *Front. Robot. AI* 3. doi:10.3389/frobt.2016.00024

Obaid, M., Johal, W., and Mubin, O. (2020). "Domestic Drones," in Proceedings of the 8th International Conference on Human-Agent Interaction, Online, November 10–13, 2020 (New York City, NY: Association for Computing Machinery, HAI'20), 196–203. doi:10.1145/3406499.3415076

Obaid, M., Sandoval, E. B., Złotowski, J., Moltchanova, E., Basedow, C. A., and Bartneck, C. (2016). "Stop! that Is Close Enough. How Body Postures Influence Human-Robot Proximity," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication, New York City, NY, August 26–31, 2016 (New York City, NY: RO-MAN), 354–361. doi:10.1109/roman.2016.7745155

Olde Keizer, R. A. C. M., van Velsen, L., Moncharmont, M., Riche, B., Ammour, N., Del Signore, S., et al. (2019). Using Socially Assistive Robots for Monitoring and Preventing Frailty Among Older Adults: A Study on Usability and User Experience Challenges. *Health Technol.* 9, 595–605. doi:10.1007/s12553-019-00320-9

Omokawa, R., Kobayashi, M., and Matsuura, S. (2019). "Expressing the Personality of a Humanoid Robot as a Talking Partner in an Elementary School Classroom," in *Universal Access in Human-Computer Interaction. Theory, Methods and Tools*. Editors M. Antona and C. Stephanidis (Cham: Springer International Publishing), 494–506. doi:10.1007/978-3-030-23560-4_36

Omokawa, R., and Matsuura, S. (2018). "Development of Thought Using a Humanoid Robot in an Elementary School Classroom," in *Universal Access in Human-Computer Interaction. Virtual, Augmented, and Intelligent Environments*. Editors M. Antona and C. Stephanidis (Cham: Springer International Publishing), 541–552. doi:10.1007/978-3-319-92052-8_43

Onnasch, L., and Roesler, E. (2020). A Taxonomy to Structure and Analyze Human-Robot Interaction. *Int. J. Soc. Robotics* 13, 833–849. doi:10.1007/s12369-020-00666-5

Pan, Y., Okada, H., Uchiyama, T., and Suzuki, K. (2013). "Direct and Indirect Social Robot Interactions in a Hotel Public Space," in 2013 IEEE International Conference on Robotics and Biomimetics, Shenzhen, China, December 12–14, 2013 (New York City, NY: ROBIO 2013), 1881–1886. doi:10.1109/ROBIO.2013.6739743

Peters, R., Broekens, J., and Neerincx, M. (2017). "Robots Educate in Style: The Effect of Context and Non-verbal Behaviour on Children's Perceptions of Warmth and Competence," in RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication, Lisbon, Portugal, August 28–September 1, 2017, 449–455. doi:10.1109/ROMAN.2017.8172341

Petrović, D., Kićinbaći, L., Petric, F., and Kovačić, Z. (2019). "Autonomous Robots as Actors in Robotics Theatre - Tribute to the Centenary of r.u.R," in In 2019 European Conference on Mobile Robots (ECMR), 1–7. doi:10.1109/ECMR.2019.8870908

Rahman, R. A. A., Hanapiah, F. A., Basri, H. H., Malik, N. A., and Yussof, H. (2015). Use of Humanoid Robot in Children with Cerebral Palsy: The Ups and Downs in Clinical Experience. *Proced. Computer Sci.* 76, 394–399. doi:10.1016/j.procs.2015.12.316

Rajamohan, V., Scully-Allison, C., Dascalu, S., and Feil-Seifer, D. (2019). "Factors Influencing the Human Preferred Interaction Distance," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication, New Delhi, India, October 14–18, 2019 (New York City, NY: RO-MAN), 1–7. doi:10.1109/RO-MAN46459.2019.8956404

Rakhymbayeva, N., Amirova, A., and Sandygulova, A. (2021). A Long-Term Engagement with a Social Robot for Autism Therapy. *Front. Robot. AI* 8, 669972. doi:10.3389/frobt.2021.669972

Ribeiro, T., Pereira, A., Di Tullio, E., and Paiva, A. (2016). "The SERA Ecosystem: Socially Expressive Robotics Architecture for Autonomous Human-Robot Interaction," in AAAI Spring Symposium - Technical Report, Palo Alto, CA, March 21–23, 2016, 155–163.

Riccio, F., Capobianco, R., and Nardi, D. (2016). "Learning Human-Robot Handovers through -STAM: Policy Improvement with Spatio-Temporal Affordance Maps," in IEEE-RAS International Conference on Humanoid Robots, Cancun, Mexico, November 15–17, 2016, 857–863. doi:10.1109/HUMANOIDS.2016.7803373

Riek, L. (2012). Wizard of Oz Studies in Hri: A Systematic Review and New Reporting Guidelines. *Jhri* 1, 119–136. doi:10.5898/jhri.1.1.riek

Robaczewski, A., Bouchard, J., Bouchard, K., and Gaboury, S. (2020). Socially Assistive Robots: The Specific Case of the Nao. *Int. J. Soc. Robotics* 13, 795–831. doi:10.1007/s12369-020-00664-7

Ros, R., Baroni, I., and Demiris, Y. (2014). Adaptive Human-Robot Interaction in Sensorimotor Task Instruction: From Human to Robot Dance Tutors. *Robotics Autonomous Syst.* 62, 707–720. doi:10.1016/j.robot.2014.03.005

Ros (2010). Robots Using Ros: Aldebaran Nao. Available at: https://www.ros.org/news/2010/03/robots-using-ros-aldebaran-nao.html (Accessed June 17, 2021).

Rosenberg-Kima, R., Koren, Y., Yachini, M., and Gordon, G. (2019). "Human-robot-collaboration (Hrc): Social Robots as Teaching Assistants for Training Activities in Small Groups," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, South Korea, March 11–14, 2019 (IEEE), 522–523. doi:10.1109/hri.2019.8673103

Roshdy, A., Karar, A. S., Al-Sabi, A., Barakeh, Z. A., El-Sayed, F., Alkork, S., et al. (2019). "Towards Human Brain Image Mapping for Emotion Digitization in Robotics," in 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART, Paris, France, April 24–26, 2019, 1–5.

Rossi, S., Dell'Aquila, E., and Bucci, B. (2019). "Evaluating the Emotional Valence of Affective Sounds for Child-Robot Interaction,". *Social Robotics*. Editors M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, Á. Castro-González, et al. (Cham: Springer International Publishing), 505–514. doi:10.1007/978-3-030-35888-4_47

Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B. W., and Picard, R. W. (2017). Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study. *Front. Robot. AI*, 4. doi:10.3389/frobt.2017.00036

Ruffin, M., Nias, J., Taylor, K., Singleton, G., and Sylvain, A. (2020). Character Development to Facilitate Retention in a Storytelling Robot. Proceedings of the 2020 ACM Southeast Conference, Tampa, FL, April 2–4, 2020. New York City, NY: ACM, 276–279. doi:10.1145/3374135.3385315

Saleh, M. A., Hanapiah, F. A., and Hashim, H. (2020). Robot Applications for Autism: A Comprehensive Review. *Disabil. Rehabil. Assist. Technol.* 16, 1–23. doi:10.1080/17483107.2019.1685016

Salter, T., Michaud, F., and Larouche, H. (2010). How Wild Is Wild? A Taxonomy to Characterize the 'wildness' of Child-Robot Interaction. *Int. J. Soc. Robotics* 2, 405–415. doi:10.1007/s12369-010-0069-4

Sandygulova, A., Johal, W., Zhexenova, Z., Tleubayev, B., Zhanatkyzy, A., Turarova, A., et al. (2020). "CoWriting Kazakh: Learning a New Script with a Robot," in ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, March 23–26, 2020, 113–120. doi:10.1145/3319502.3374813

Sandygulova, A., and O'Hare, G. M. P. (2015). "Children's Perception of Synthesized Voice: Robot's Gender, Age and Accent," in *Social Robotics*. Editors A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi (Cham: Springer International Publishing), 594–602. doi:10.1007/978-3-319-25554-5_59

Sarabia, M., Young, N., Canavan, K., Edginton, T., Demiris, Y., and Vizcaychipi, M. P. (2018). Assistive Robotic Technology to Combat Social Isolation in Acute Hospital Settings. *Int. J. Soc. Robotics* 10, 607–620. doi:10.1007/s12369-017-0421-z

Shamsuddin, S., Ismail, L. I., Yussof, H., Ismarrubie Zahari, N., Bahari, S., Hashim, H., et al. (2011). "Humanoid Robot Nao: Review of Control and Motion Exploration," in 2011 IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia, November 25–27, 2011, 511–516. doi:10.1109/ICCSCE.2011.6190579

Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., and Zahari, N. I. (2012). Humanoid Robot NAO Interacting with Autistic Children of Moderately Impaired Intelligence to Augment Communication Skills. *Proced. Eng.* 41, 1533–1538. doi:10.1016/j.proeng.2012.07.346

Shidujaman, M., and Mi, H. (2018). ""Which Country Are You from?" A Cross-Cultural Study on Greeting Interaction Design for Social Robots," in *Cross-Cultural Design. Methods, Tools, and Users*. Editor P.-L. P. Rau (Cham: Springer International Publishing), 362–374. doi:10.1007/978-3-319-92141-9_28

Sinnema, L., and Alimardani, M. (2019). "The Attitude of Elderly and Young Adults towards a Humanoid Robot as a Facilitator for Social Interaction," in *Social Robotics*. Editors M. A. Salichs, S. S. Ge, E. I. Barakova, J.-J. Cabibihan, A. R. Wagner, Á. Castro-González, et al. (Cham: Springer International Publishing), 24–33. doi:10.1007/978-3-030-35888-4_3

Sorce, M., Pointeau, G., Petit, M., Mealier, A.-L., Gibert, G., and Dominey, P. F. (2015). "Proof of Concept for a User-Centered System for Sharing Cooperative Plan Knowledge over Extended Periods and Crew Changes in Space-Flight Operations," in Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, Kobe, Japan, August 31–September 4, 2015, 776–783. doi:10.1109/ROMAN.2015.7333565

Stadler, S., Weiss, A., and Tscheligi, M. (2014). "I Trained This Robot: The Impact of Pre-experience and Execution Behavior on Robot Teachers," in The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, August 25–29, 2014 (New York City, NY: RO-MAN), 1030–1036. doi:10.1109/ROMAN.2014.6926388

Stankevich, L., and Sonkin, K. (2016). "Human-robot Interaction Using Brain-Computer Interface Based on Eeg Signal Decoding," in *Interactive Collaborative Robotics*. Editors A. Ronzhin, G. Rigoll, and R. Meshcheryakov (Cham: Springer International Publishing), 99–106. doi:10.1007/978-3-319-43955-6_13

Stojanovska, F., Toshevska, M., Kirandziska, V., and Ackovska, N. (2018). Emotion-Aware Teaching Robot: Learning to Adjust to User's Emotional State. *ICT Innov.* 940, 59–74. doi:10.1007/978-3-030-00825-3_6

Sturgeon, S., Palmer, A., Blankenburg, J., and Feil-Seifer, D. (2019). "Perception of Social Intelligence in Robots Performing False-Belief Tasks," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication, New Delhi, India, October 14–18, 2019 (New York City, NY: RO-MAN), 1–7. doi:10.1109/RO-MAN46459.2019.8956467

Sucharew, H., and Macaluso, M. (2019). Methods for Research Evidence Synthesis: The Scoping Review Approach. *J. Hosp. Med.* 14, 416. doi:10.12788/jhm.3248

Tapus, A., Peca, A., Aly, A., A. Pop, C., Jisa, L., Pintea, S., et al. (2012). Children with Autism Social Engagement in Interaction with Nao, an Imitative Robot. *Is* 13, 315–347. doi:10.1075/is.13.3.01tap

Tariq, S., Baber, S., Ashfaq, A., Ayaz, Y., Naveed, M., and Mohsin, S. (2016). "Interactive Therapy Approach through Collaborative Physical Play between a Socially Assistive Humanoid Robot and Children with Autism Spectrum Disorder," in *Social Robotics*. Editors A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He (Cham: Springer International Publishing), 561–570. doi:10.1007/978-3-319-47437-3_55

Thobbi, A., and Sheng, W. (2010). "Imitation Learning of Hand Gestures and its Evaluation for Humanoid Robots," in The 2010 IEEE International Conference on Information and Automation, Harbin, China, June 20–23, 2010 (IEEE), 60–65. doi:10.1109/ICINFA.2010.5512333

Tielman, M., Neerincx, M., Meyer, J.-J., and Looije, R. (2014). "Adaptive Emotional Expression in Robot-Child Interaction," in ACM/IEEE International Conference on Human-Robot Interaction, Bielefeld, Germany, March 3–6, 2014, 407–414. doi:10.1145/2559636.2559663

Tozadore, D., Pinto, A., Romero, R., and Trovato, G. (2017). "Wizard of Oz vs Autonomous: Children's Perception Changes According to Robot's Operation Condition," in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication, Lisbon, Portugal, August 28–September 1, 2017 (New York City, NY: RO-MAN), 664–669. doi:10.1109/ROMAN.2017.8172374

Tozadore, D., Pinto, A., and Romero, R. (2016). "Variation in a Humanoid Robot Behavior to Analyse Interaction Quality in Pedagogical Sessions with Children," in Proceedings - 13th Latin American Robotics Symposium and 4th Brazilian Symposium on Robotics, Recife, Brazil, October 8–12, 2016 (New York City, NY: LARS/SBR), 133–138. doi:10.1109/lars-sbr.2016.29

Tsiakas, K., Kyrarini, M., Karkaletsis, V., Makedon, F., and Korn, O. (2018). A Taxonomy in Robot-Assisted Training: Current Trends, Needs and Challenges. *Technologies* 6, 119. doi:10.3390/technologies6040119

Turp, M., González, J. C., Pulido, J. C., and Fernández, F. (2019). Developing a Robot-Guided Interactive Simon Game for Physical and Cognitive Training. *Int. J. Hum. Robot.* 16, 1950003. doi:10.1142/s0219843619500038

van der Woerdt, S., and Haselager, P. (2019). When Robots Appear to Have a Mind: The Human Perception of Machine agency and Responsibility. *New Ideas Psychol.* 54, 93–100. doi:10.1016/j.newideapsych.2017.11.001

VanderPlas, J., Granger, B., Heer, J., Moritz, D., Wongsuphasawat, K., Satyanarayan, A., et al. (2018). Altair: Interactive Statistical Visualizations for python. *Joss* 3, 1057. doi:10.21105/joss.01057

Vircikova, M., and Sincak, P. (2013). Experience with the Children-Humanoid Interaction in Rehabilitation Therapy for Spinal Disorders. *Adv. Intell. Syst. Comput.* 208, 347–357. doi:10.1007/978-3-642-37374-9_34

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). "Second Language Tutoring Using Social Robots: a Large-Scale Study," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, South Korea, March 11–14, 2019 (IEEE), 497–505. doi:10.1109/hri.2019.8673077

Wang, X., Williams, M.-A., Gärdenfors, P., Vitale, J., Abidi, S., Johnston, B., et al. (2014). "Directing Human Attention with Pointing," in IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions, Edinburgh, UK, August 25–29, 2014, 174–179. doi:10.1109/roman.2014.6926249

Wei, Y., and Zhao, J. (2016). Designing Robot Behavior in Human Robot Interaction Based on Emotion Expression. *Ir* 43, 380–389. doi:10.1108/IR-08-2015-0164

Werner, F., Krainer, D., Oberzaucher, J., and Werner, K. (2013). Evaluation of the Acceptance of a Social Assistive Robot for Physical Training Support Together with Older Users and Domain Experts. *Assistive Technology Res. Ser.* 33, 137–142. doi:10.3233/978-1-61499-304-9-137

Wong, C. J., Ling Tay, Y., Lew, L. W. C., Koh, H. F., Xiong, Y., and Wu, Y. (2018). "Advbot: Towards Understanding Human Preference in a Human-Robot Interaction Scenario," in 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, November 18–21, 2018 (IEEE), 1305–1309. doi:10.1109/ICARCV.2018.8581325

Wood, L. J., Zaraki, A., Robins, B., and Dautenhahn, K. (2019). Developing Kaspar: A Humanoid Robot for Children with Autism. *Int. J. Soc. Robotics* 13, 491–508. doi:10.1007/s12369-019-00563-6

Wu, Y., Wang, R., Tay, Y. L., and Wong, C. J. (2018). "Investigation on the Roles of Human and Robot in Collaborative Storytelling," in Proceedings - 9th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kula Lumpur, Malaysia, December 12–15, 2017 (New York City, NY: APSIPA ASC), 63–68. doi:10.1109/APSIPA.2017.8282003

Xu, J., Broekens, J., Hindriks, K., and Neerincx, M. A. (2015). Mood Contagion of Robot Body Language in Human Robot Interaction. *Auton. Agent Multi-agent Syst.* 29, 1216–1248. doi:10.1007/s10458-015-9307-3

Yadollahi, E., Johal, W., Paiva, A., and Dillenbourg, P. (2018). "When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration," in Proceedings of the 17th ACM Conference on Interaction Design and Children, Trondheim, Norway, June 17–24, 2018, 195–206. doi:10.1145/3202185.3202743

Yanco, H. A., and Drury, J. L. (2004). "Classifying Human-Robot Interaction: an Updated Taxonomy," in 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), The Hague, Netherlands, October 10–13, 2004, 2841–2846.

Ye, P., Wang, T., and Wang, F.-Y. (2018). A Survey of Cognitive Architectures in the Past 20 Years. *IEEE Trans. Cybern.* 48, 3280–3290. doi:10.1109/tcyb.2018.2857704

Zhexenova, Z., Amirova, A., Abdikarimova, M., Kudaibergenov, K., Baimakhan, N., Tleubayev, B., et al. (2020). A Comparison of Social Robot to Tablet and Teacher in a New Script Learning Context. *Front. Robot AI* 7, 99. doi:10.3389/frobt.2020.00099

Złotowski, J., Strasser, E., and Bartneck, C. (2014). "Dimensions of Anthropomorphism," in ACM/IEEE International Conference on Human-Robot Interaction, Biefeld, Germany, March 3–6, 2014, 66–73. doi:10.1145/2559636.2559679

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

METHODS
published: 03 December 2021
doi: 10.3389/frobt.2021.704119

LEADOR: A Method for End-To-End Participatory Design of Autonomous Social Robots

Katie Winkle[1*†], *Emmanuel Senft*[2*†] *and Séverin Lemaignan*[3]

[1]*Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden,* [2]*Department of Computer Sciences, University of Wisconsin–Madison, Madison, WI, United States,* [3]*Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom*

OPEN ACCESS

Edited by:
Maartje M. A. De Graaf, Utrecht University, Netherlands

Reviewed by:
Vincenzo D'Andrea, University of Trento, Italy
John Sören Pettersson, Karlstad University, Sweden

***Correspondence:**
Katie Winkle
winkle@kth.se
Emmanuel Senft
esenft@wisc.edu

[†]*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *01 May 2021*
Accepted: *08 October 2021*
Published: *03 December 2021*

Citation:
Winkle K, Senft E and Lemaignan S (2021) LEADOR: A Method for End-To-End Participatory Design of Autonomous Social Robots. Front. Robot. AI 8:704119. doi: 10.3389/frobt.2021.704119

Participatory design (PD) has been used to good success in human-robot interaction (HRI) but typically remains limited to the early phases of development, with subsequent robot behaviours then being hardcoded by engineers or utilised in Wizard-of-Oz (WoZ) systems that rarely achieve autonomy. In this article, we present LEADOR (Led-by-Experts Automation and Design Of Robots), an *end-to-end* PD methodology for domain expert co-design, automation, and evaluation of social robot behaviour. This method starts with typical PD, working with the domain expert(s) to co-design the interaction specifications and state and action space of the robot. It then replaces the traditional offline programming or WoZ phase by an *in situ* and online teaching phase where the domain expert can live-program or teach the robot how to behave whilst being embedded in the interaction context. We point out that this live teaching phase can be best achieved by adding a learning component to a WoZ setup, which captures implicit knowledge of experts, as they intuitively respond to the dynamics of the situation. The robot then progressively learns an appropriate, expert-approved policy, ultimately leading to full autonomy, even in sensitive and/or ill-defined environments. However, LEADOR is agnostic to the exact technical approach used to facilitate this learning process. The extensive inclusion of the domain expert(s) in robot design represents established responsible innovation practice, lending credibility to the system both during the teaching phase and when operating autonomously. The combination of this expert inclusion with the focus on *in situ* development also means that LEADOR supports a mutual shaping approach to social robotics. We draw on two previously published, foundational works from which this (generalisable) methodology has been derived to demonstrate the feasibility and worth of this approach, provide concrete examples in its application, and identify limitations and opportunities when applying this framework in new environments.

Keywords: social robotics, HRI, participatory design, mutual shaping of technology and society, autonomous robots, robot development

1 INTRODUCTION

In the context of robotics research, participatory design (PD) attempts to empower non-roboticists such that they can shape the direction of robotics research and actively collaborate in robot design (Lee et al., 2017). Typically, PD is achieved by researchers running workshops or focus groups with end users and/or domain experts. Output may include potential use case scenarios (Jenkins and Draper, 2015), design guidelines/recommendations (Winkle et al., 2018), and/or prototype robot behaviours (Azenkot et al., 2016). Šabanović identified such methods as appropriate for the pursuit of a mutual shaping approach in robot design that is one that recognises the dynamic interactions between social robots and their context of use (Šabanović, 2010), an approach that we find compelling for designing effective and acceptable social robots efficiently. However, the automation of social robot behaviour, which requires a significant technical understanding of robotics and artificial intelligence (AI), is not typically considered during such activities.

Instead, common methods for the automation of social robot behaviour include utilising models based on human psychology (e.g., Theory of Mind, Lemaignan et al., 2017) or animal behaviour (Arkin et al., 2001) or attempting to observe and replicate human-human interaction behaviours (e.g., Sussenbach et al., 2014). This limits the potential for direct input from domain experts (teachers, therapists, etc.) who are skilled in the use of social interaction in complex scenarios. Previous work with such experts has demonstrated that a lot of the related expertise is intuitive and intangible, making it difficult to access in a way that can easily inform robot automation (Winkle et al., 2018). This is somewhat addressed by methods that capture domain expert operation of a robot directly, for example, end user programming tools (e.g., Leonardi et al., 2019) or learning from expert teleoperation of robots (e.g., Sequeira et al., 2016). However, these methods tend to focus on offline learning/programming. As such, there is no opportunity for experts to create an adequate, situated mental model of the capabilities of the robot, limiting the guarantee of appropriate behaviour when the robot is eventually deployed to interact with users autonomously.

Instead, we argue that robots should be automated by domain experts themselves, in real time, and whilst being situated in the interaction context; and that this automation should be done through a direct, bi-directional interaction between the expert and the robot. We refer to this as the *teaching phase*, where the robot is taught what to do by the domain expert, regardless of whether it is, e.g., a machine learning algorithm or an authoring tool that underpins this interaction. This live, *in situ*, and interactive approach allows *mutual shaping* to occur during robot automation, as the expert defines the program of the robot in response to the evolving dynamics of the social context into which the robot has been deployed.

1.1 Supporting a Mutual Shaping Approach to Robot Design

Šabanović proposed a *mutual shaping* approach to social robot design that is one that recognises the dynamic interactions between social robots and their context of use, in response to their finding that most roboticists were taking a technologically deterministic view of the interaction between robots and society (Šabanović, 2010). Studies of real-world human-robot interaction (HRI) motivate such an approach, because they demonstrate how mutual shaping effects impact robot effectiveness upon deployment in the real world. For example, the use and acceptance of robots in older adult health settings has been shown to be affected by situation and context of use factors such as user age and gender, household type, and the prompting of its use by others (Chang and Šabanović, 2015; de Graaf et al., 2015), i.e., factors unrelated to the functionality of the robot. The pursuit of a mutual shaping approach, primarily through use of PD and in-the-wild robot evaluation methods, gives the best possible chance of identifying and accounting for such factors during the design and development process, such that the robot has maximum positive impact on its eventual long-term deployment.

To this end, Šabanović describes four key practices that underpin a mutual shaping approach to support a "*socially robust understanding of technological development that enables the participation of multiple stakeholders and disciplines*":

1) Evaluating robots in society: HRI studies and robot evaluations should be conducted "in the wild", i.e., in the environments and context of use for which they are ultimately intended to be deployed (Ros et al., 2011).
2) Studying socio-technological ecologies: Robot design should be informed by systematic study of the context of use, and evaluation of robots should consider impact on the socio-technology ecology into which they have been deployed.
3) Outside-in design: Design constraints should be defined by empirical social research and the social context of use, rather than technical capabilities, and evaluation should be based on user experiences rather than internal measures of technical capability.
4) Designing with users: Stakeholders (those who will be directly affected by the deployment and use of the robot) should be included in identifying robot applications and thinking about how robots will be used and in designing the robot and its behaviour(s).

However, as we explain in **Section 2**, robot development at present typically represents a discontinuous process, particularly broken up by the automation of social robot behaviour. It still tends to heavily rely on technical expertise, executed in research/development environments rather than the real world, with little active inclusion of domain experts or other expert stakeholders. This discontinuity also represents a key hurdle to truly multi-disciplinary working, a disconnection between those of different academic backgrounds on the research team, which can result in a number of practical challenges and frustrations.

1.2 The *Led-By-Experts Automation and Design of Robots* Method

The generalisable method that we provide in this work is derived from two (independently undertaken) foundational works. First is the educational robot by Senft et al. (2019) for school children, in which a psychologist taught a robot to support children in an

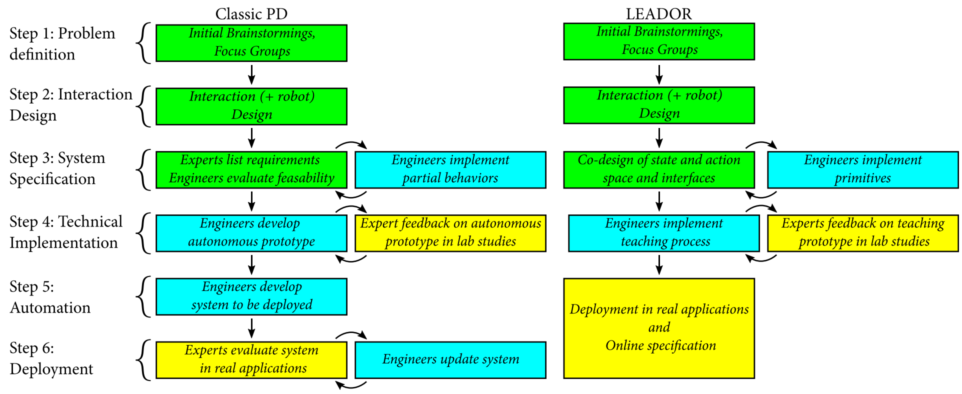


FIGURE 1 | Comparison between a classic participatory design (PD) approach and LEADOR, our proposed end-to-end participatory design approach. Green activities represent joint work between domain experts, multidisciplinary researchers, and/or engineers; yellow activities are domain expert-led; blue activities are engineer-led. Compared to typical PD, the two key differences in our approach are the focus on developing a *teaching system* instead of a *final autonomous behaviour* in step 4, and the combining of autonomous action policy definition and deployment in the real world into a single step 5 + 6. In addition, our method reduces the amount of work that is carried out independently by engineers (i.e., with no domain expert or non-roboticist input).

educational activity. After the teaching phase with 25 children, the robot was evaluated in further autonomous interaction with children, which demonstrated the opportunity of online teaching as a way to define autonomous robot behaviours.

Second is the robot fitness coach by Winkle et al. (2020). This work built upon the work by Senft et al. (2019) by integrating the online teaching method into an end-to-end PD process, whereby the same professional fitness instructor was involved in the co-design, automation, and evaluation of a robot fitness coach. This work also demonstrated the value of online teaching when compared to expert-designed heuristics as a next best alternative for defining autonomous robot behaviours with domain expert involvement. Both studies used a teaching phase where a domain expert interacted with the robot to create an interactive behaviour, and in both studies, the resulting autonomous robot behaviour was evaluated with success.

From these works, we have derived a five-step, generalisable method for end-to-end PD of autonomous social robots (*Led-by-Experts Automation and Design Of Robots* or LEADOR), depicted alongside typical PD in **Figure 1**. The key stages of our approach, as referenced in the figure, can be summarised as follows:

1) Problem definition: *Initial brainstorming, studies of context of use, and studies with stakeholders.*
2) Interaction design: *Detailed refinement of robot application and interaction scenario, and choice/design of robot platform.*
3) System specification: *Co-design of the action space of the robot, input space, and teaching interface.*
4) Technical implementation: *Realisation of the third stage through technical implementation of underlying architecture and all sub-components and tools required for the teaching phase.*
5) Real-world deployment: *Robot is deployed in the real world, where a teaching phase is undertaken, led by the domain expert(s), to create autonomous robot behaviour.*

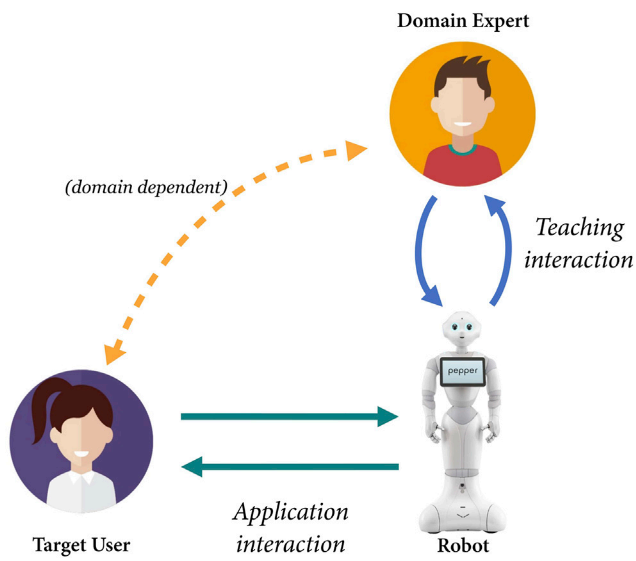


FIGURE 2 | Three-way interaction between the domain expert, the robot, and the target user through which the expert teaches the robot during a teaching phase upon real-world deployment. Robot automation is therefore happening in the real world, whereas the robot is fully embedded in its long-term application context. The expert is teaching the robot through bi-directional communication, as the robot interacts with the target user. The extent of interaction(s) between the domain expert and target user should be consistent with what is envisaged for long-term deployment of the robot and is domain-dependent. People vector created by studiogstock - www.freepik.com.

The cornerstone of our method is to facilitate robot automation through direct interaction between the expert and the robot, during a "teaching phase", whereby the domain expert teaches the robot what to do during real interaction(s) with the target user. The resultant interaction is depicted in **Figure 2**. Regardless of the specifics of the final interaction, the output of this phase is a robot that *can* operate autonomously but could also allow for continued expert-in-the-loop supervision and/or behaviour correction/additional training.

Through our foundational works, we demonstrate the flexibility in our method for developing autonomous robots for different long-term interaction settings. The educational robot by Senft et al. (2019) was intended for diadic, unsupervised robot-user interactions, whereas the robot fitness coach by Winkle et al. (2020) was intended primarily for diadic robot-user interactions but to be complimented with additional expert-user interactions/supervision and with additional expert-robot-user teaching interactions if necessary. LEADOR could also be used to design robots with other interaction requirements, e.g., an autonomous robot to be used in fully triadic expert-robot-user interactions or to facilitate permanent expert supervision and validation of autonomous behaviour.

In this paper, we have combined our experiences from these foundational works to propose an end-to-end PD process, centred around an *in situ* teaching phase, that uniquely delivers on the promises of mutual shaping and PD. We suggest that this approach is as *practical* as it is *responsible*, because our foundational studies demonstrate that we were able to create appropriate, intelligent, and autonomous social robot behaviour for complex application environments in a timely manner. As detailed in Senft et al. (2015) and Senft et al. (2019), this teaching phase is achieved by deploying the robot in the proposed use case and it is initially controlled completely by a human "teacher". The teacher can progressively improve the robot behaviour *in situ* and generate a mental model of the policy of the robot. This teaching can continue until the domain expert is confident that the robot can satisfactorily operate autonomously. This approach therefore allows non-roboticist, domain experts to actively participate in creating autonomous robot behaviour. It also allows for the continual shaping of robot behaviour, because teaching can be seamlessly (re-)continued at any time to address any changes in the interaction dynamics, therefore better supporting a mutual shaping approach. We suggest that our methodology is particularly appropriate for use cases, in which difficult-to-automate and/or difficult-to-explain "intuitive" human domain expertise and experience are needed to inform personalised interaction and engagement (e.g., socially assistive robotics). The result then is an autonomous robot that has been designed, developed, and evaluated (by a multi-disciplinary research team) directly in conjunction with domain experts, within its real-world context of use, that can intelligently respond to complex social dynamics in ways that would have otherwise been very difficult to automate.

For clarity, hereafter, we use the term *domain expert* (or *teacher*) to refer to experts in an application domain. For example, these domain experts could be therapists, shop owners, or school teachers. These experts interact with the robot and specify its behaviour in a *teaching interaction* (even if no actual machine learning might be involved). On the other hand, *engineers* or developers refer to people with technical expertise in robotics or programming. They are the ones typically programming a robot behaviour or developing tools to be used by domain experts. Finally, the *target user* is the person a robot would interact with in the *application interaction*. For example, such target users could be children during a therapy session or store customers in a shopping interaction. This population is expected to interact with the robot at the point of use, rather than be the ones directly defining the autonomous robot behaviour.

2 RELATED WORK

2.1 Participatory Design

PD is fundamentally concerned with involving the people who will use and/or will be affected by a technology in the design of that technology, with a focus on mutual learning between participants who typically represent either *domain experts* (users) or *technology experts* (designers) (Simonsen and Robertson, 2012). Contemporary PD has been concerned with combining typical, iterative PD practices with evaluation of the design in use under the concept of *sustained PD* or design as "emerging change" (Simonsen et al., 2010; Simonsen et al., 2014). Originally posed in the context of large-scale information systems projects, the sustained PD approach presented by Simonsen and Hertzum (2012) not only emphasises the evaluation of systems by exposing them to real-situated work practices but also notes associated challenges regarding management of a stepwise implementation process and the conducting of realistic, large-scale PD experiments.

This focus on implementation of the new technology as part of PD also therefore raises the notion of user participation in that implementation process. Fleron et al. (2012) demonstrated this in the context of first designing an electronic whiteboard with healthcare practitioners, followed by healthcare practitioner-led implementation of that whiteboard at two emergency departments. Notably, this work investigated differences in the experiences between staff who did and did not participate in this implementation process. Specifically, the participating staff (those responsible for the system implementation) identified some difficulties in understanding their role/responsibilities. The non-participating staff expressed a desire for earlier (pre-deployment) testing of the system but demonstrated positive "buy-in" of the system, nonetheless, with the authors positing this linked to reputed credibility of the system given it was (co-)developed and implemented by their peers. This points towards not only the potential benefits but also the challenges when trying to include end users and/or domain experts in this sustained PD process.

Specifically concerning PD, AI, and machine learning, Bratteteig and Verne (2018) note three key challenges when using PD with/for AI systems. First is that users and designers might struggle to conceptualise the possibilities and limitation of AI; second is that (machine learning–based) AI systems develop over time and hence are difficult to evaluate within a typical PD experimental period; and third is how to distinguish between "normal" use and training. Overall, the authors conclude that AI and machine learning can be part of a PD process, but that AI poses complex challenged that go beyond the scope of typical PD projects.

From the PD literature then, there is a clear motivation to explore PD processes that go beyond initial design to also include implementation and to understand how best to approach the PD of machine learning–based AI systems. The notion of using expert-in-the-loop machine learning for sustained PD that also includes implementation specifically sits at this exact intersection. However, whilst many contemporary PD works have described applications in the development and implementation of information technology systems, there seem to be very few (if any) that consider autonomous, social robot design, and development. Considering literature from the HRI field, however, a number of (interdisciplinary) HRI researchers have utilised and drawn from PD in the context of designing robots and their applications.

Participatory Design in Social Robotics

Here, we identify relevant works utilising PD and other related methodologies specifically in the context of designing social HRI. Notably, these works primarily originate from the HRI community, as opposed to the PD community, but most such works showcase one (or more) of the following methodologies:

1) *Ethnographic/"In-the-Wild" Studies*: These typically focus on understanding situated use and/or emergent behaviour(s) on deployment of a robot into the real world. Concerning robot design, such studies are inherently limited to the testing of prototypes, Wizard-of-Oz (WoZ) systems, or finished products (e.g., Forlizzi, 2007 and Chang and Šabanović, 2015). However, they might be used to inform initial design requirements (and their iteration) through observation of the use case environment and user behaviour.
2) *User-Centred Design (UCD)*: This aims to understand and incorporate user perspective and needs into robot design. Typically, researchers set the research agenda based on prior assumptions regarding the context of use and proposed robot application (e.g., Louie et al., 2014, Wu et al., 2012, and Beer et al., 2012).
3) *Participatory Design (PD)*: This encourages participants (users, stakeholders, etc.) to actively join in decision making processes which shape robot design and/or the direction of research. This typically involves participants having equal authority as the researchers and designers, with both engaging in a two-way exchange of knowledge and ideas (e.g., Azenkot et al., 2016 and Björling and Rose, 2019).

Lee et al. give a good overview of the above practices as employed in social robot and HRI design/research, with a particular focus on how the shortcomings of 1 and 2 can be addressed using PD (Lee et al., 2017). The authors use a case study of social robot PD from their own work to highlight a number of PD design principles for informing social robot design and further development of PD methodologies. They particularly highlight the empowering of PD participants to become active "robot co-designers" through *mutual learning*, as introduced previously, whereby there is a two-way exchange of knowledge and experience between researchers/designers and expert stakeholders. Through this process, users learn about, e.g., robot capabilities, such that they are better informed to contribute to discussions on potential applications, whilst the researchers/designers come to learn more about the realities of the proposed context of use from the perspective of the users.

Since publication of work of Lee et al., PD methods have been gaining visibility for the design of social robots, with other roboticists further refining PD methods and best practice for their use in social robotics and HRI. As such, PD works relating to ours can be grouped into two categories:

1) Results-focused publications that utilised PD methods.
2) Methodology-focused publications in which the authors share or reflect on PD methods for use in social robot/HRI design.

Works on 1) have typically taken the form of researchers working closely with prospective users and/or other stakeholders *via* focus groups, interviews, workshops, etc., with the researchers then concatenating their results to produce potential use case scenarios (Jenkins and Draper, 2015), design guidelines/recommendations (Winkle et al., 2018), and/or prototype robot behaviours (Azenkot et al., 2016). For example, Azenkot et al. (2016) used PD to generate specifications for a socially assistive robot for guiding blind people through a building. The study of the authors consisted of multiple sessions including interviews, a group workshop, and individual user-robot prototyping sessions. The initial interviews were used, in part, to brief participants about robot capabilities. The group session was used to develop a conceptual storyboard of robot use, identifying interactions between the robot guide and the user.

Winkle et al. (2018) conducted a study with therapists, utilising a novel focus group methodology combined with follow-up individual interviews to generate an expert-informed set of design guidelines for informing the design and use of socially assistive robots in rehabilitative therapies. The topic guides for each part of the study were designed to help the researchers to understand typical therapy practice and therapist best practices for improving patient engagement and to explore the ideas and opinions of the therapists on the potential role(s) social robots, which might play in rehabilitation. A key finding of this work was the extent to which intuitive, instantaneous behaviour of therapists is driven by situational factors specific to each individual client, making it difficult, for example, to extract any clear cut heuristics that might inform generalisable, autonomous social robot behaviour directly. The resultant design guidelines therefore suggested that socially assistive robots require "high-level" personalisation to each user as well as the ability to adapt, in real time, to, e.g., the performance of the user and other situational factors. This is one of the key works that motivates our effort to therefore facilitate expert-led, *in situ* robot teaching, to capture this sort of tacit social intelligence.

A follow up publication by the same authors then comes under category 2). Specifically, the authors provide more detail on their focus group methodology, and how it reflects a mutual shaping approach to social robot design, alongside a general guide in how it might be applied to other domains (Winkle et al., 2019a). The method combines elements of PD and UCD and utilises a demonstration of robot capabilities to support mutual learning between the researchers and participants. To evidence how this method supported mutual shaping in their work and why this was beneficial, the authors identify specific project-related considerations as well as new research directions that could only be identified in conjunction with their domain expert participants and also note that taking part in a focus group significantly (positively) impacted on the acceptance of participants of social robots.

Further, under category 2), Björling and Rose (2019) shared PD methods that they used in the context of taking an overall *human-centred design* approach to co-designing robots for improved mental health with teens. They present three method cases that cover novel and creative participatory techniques such as research design, script-writing and prototyping, and concluding with a set of PD principles for guiding design work with vulnerable participants in a human-centred domain. One of their methods revolved around inviting teens to act as WoZ robot operators. Specifically, their setup had one teen teleoperating a robot, whereas another teen recounted a (pre-scripted) stressful experience. In the second experiment, they utilised virtual reality (VR) such that one teen interacted, in an immersive VR environment, with a robot avatar teleoperated by a teen outside of that VR environment. From this study, the authors gathered data about the way teens collaborate and their perceptions of robot roles and behaviours. To this end, they demonstrated the value in expert (user) teleoperation of a proposed robot, not only to better understand both the use case requirements and user needs but also to generate exemplars of desirable autonomous robot behaviour. Alves-Oliveira et al. (2017) also demonstrated a similar use of puppeteering and role-play methods as part of a co-design study with children.

In summary, the work to date has demonstrated how PD methods can be used to study a proposed application domain in an attempt to ensure that researchers thoroughly understand the context of use and to elicit some expert knowledge for informing robot design and automation. This goes some way to supporting a mutual shaping and responsible robotics approach to social robot development. However, there remains two key disconnects in delivering truly end-to-end PD and mutual shaping in development of an autonomous social robot. First, robot automation is informed but not controlled or developed by domain experts. Second, there is a disconnection between this program definition and the real-world interaction requirements and situational specificities that will likely be crucial to overall robot success when deployed in real-world interaction.

2.2 Alternative Methods to Capture Domain Expert Knowledge

One of the key assumptions of PD in the context of robotics research is that the knowledge of the desired robot behaviour is held by domain experts and needs to be translated into programs. Typically, this translation is made by engineers, obtaining a number of heuristics from the domain experts and consequently automating the robot. Although widely applied even in PD research, this method only partially delivers on the promises of PD, because domains experts are used to inform robot behaviours but still rely on external actors (the engineers) to transform their intuition, knowledge, and expertise into actual code. Furthermore, this process can lead to a number of communication issues because members from different communities have different ways of expressing needs and desires. Nevertheless, there exist a number of alternative solutions to capture domain expert knowledge that could support a PD approach to robot automation.

2.2.1 End User Programming Tools

A first solution is to create tools to allow domain experts to create robot behaviours themselves. The research on end user development or end user programming explores tools to allow domain experts or end users to create programs without requiring coding knowledge. Typical applications are home automation, application synchronisation (e.g., IFTTT or Microsoft Flow), or video games development. In addition, end user programming has seen large interest in robotics, for example, to create autonomous robot behaviours for both industrial robots (Paxton et al., 2017; Gao and Huang, 2019) and social robots (Leonardi et al., 2019; Louie and Nejat, 2020). These *authoring* tools are often developed by engineers and then provided to users to create their own applications without relying on text-based coding, for example, by using visual or block programming (Huang and Cakmak, 2017), tangible interfaces (Porfirio et al., 2021), or augmented reality (Cao et al., 2019).

However, whilst being more friendly for users, such methods still suffer from two main drawbacks. First, the interface is often developed by engineers without necessarily following principles of PD. Second, these methods often see the programming process as a discrete step leading to a static autonomous behaviour with limited opportunity to update the robot behaviour or little focus on testing and evaluating the created behaviour in real interactions. More precisely, users of these tools might not be the actual target of the application interaction and would program robots outside of the real context of use, forcing the aspiring developers to rely on their internal simulation of how the interaction should happen. For example, a shop owner could use an authoring tool to create a welcoming behaviour for a social robot and test it on themselves whilst developing the behaviour and then deploying it on real clients with limited safeguards. In such process, the developers have to use their best guess to figure out how people might interact with the robot and often have issues to infuse the robot with tacit knowledge, such as timing for actions or proxemics. This disconnect can lead to suboptimal robot behaviour as the robot will face edge cases in the real world that the designer might not have anticipated.

2.2.2 Learning From Real-World Interaction(s)

A method to address this gap between an offline design phase and the real world is to mimic the expert whilst they perform the interaction. Using machine learning, systems can learn from the experts how robots should behave. For example, Liu et al. (2016) asked participants to role play an interaction between a shopkeeper and a client and recorded data about this interaction (e.g., location or speech of participants). From these recordings, Liu et al. learned a model of the shopkeeper, transferred it to the robot, and evaluated its HRIs. Similarly, Porfirio et al. (2019) recorded interaction traces between human actors and formalised them into finite state machine to create a robot behaviour. Whilst relying on simulated interactions, these methods provide more opportunities to developers to explore situations outside of their initial imagination.

One assumption of these methods is that robots should replicate human behaviour. Consequently, such methods allow the capture of implicit behaviours such as the timing and idiosyncrasies of human interactions. However, real-world interactions with robots might follow social norms different from ones between humans only. Consequently, learning directly from human-human interaction also presents limitations.

WoZ is an interaction paradigm widely used in robotics (Riek, 2012), whereby a robot is controlled by an expert deciding on what actions the robot should execute and when. The main advantage of this paradigm is to ensure that the robot behaviour is, at all times, appropriate to the current interaction. For this reason, WoZ has been extensively used in robot-assisted therapy and exploratory studies to explore how humans react to robots. Recent research has explored how this interaction can be used to collect data from real HRI and learn an appropriate robot behaviour. Knox et al. (2014) proposed the "Learning from the Wizard" paradigm, whereby a robot would first be controlled in a WoZ experiment used to acquire the demonstrations and then machine learning would be applied offline to define a policy. Sequeira et al. (2016) extended and applied this Learning from Demonstration (LfD), with an emphasis on the concept of "restricted-perception WoZ", in which the wizard only has access to the same input space as the learning algorithm, thus reducing the problem of correspondence between the state and action spaces used by the wizard and the ones available to the robot controller. Both of these works could support a PD approach to robot automation, because they could be used to generate an autonomous robot action policy based on data from (non-roboticist) domain expert WoZ interactions in real-world environments. Nevertheless, the typical WoZ puppeteering setup results in an absence of interaction between the design/development team and the robot, which prevents designers from having a realistic mental model of the robot behaviour and does not allow for any mutual shaping between the wizard, the robot, and the contextual environment. Traditionally, LfD separates data collecting and learning into distinct steps, limiting the opportunity to know during the teleoperated data collection process at what point "enough" training data has actually been collected, because the system can only be evaluated once the learning process is complete. Similarly, when using end user programming methods, there is little opportunity to know how the system would actually behave when deployed in the real world. This lack of knowledge about the actual robot behaviour implies that robots have to be deployed to interact in the real world with limited guaranties or safeguards ensuring their behaviours are actually efficient in the desired interaction.

2.2.3 Interactive Machine Learning

Interactive Machine Learning (IML) refers to a system learning online whilst it is being used (Fails and Olsen, 2003; Amershi et al., 2014). The premise of IML is to empower end users whilst reducing the iteration time between subsequent improvements of a learning system. Using IML to create robot behaviours through an interaction between a designer and an autonomous agent allows for full utilisation of the teaching skills of the expert. It has been shown that humans are skilled teachers who can react to the current performance of a learner and provide information specifically relevant to them (Bloom, 1984). Similarly, the previous research showed that this effect also exists, to a certain extent, when teaching robots. Using Socially Guided Machine Learning (Thomaz and Breazeal, 2008), a human teacher adapts their teaching strategy to robot behaviour and thus helps it to learn better. If able to observe (and correct) the autonomous behaviour of the robot, seeing the result of the robot behaviour as it progresses, then the expert can create a model of the knowledge, capabilities, and limitations of the robot. This understanding of the robot reduces the risk of over-trusting (both during training and/or autonomous operation) and introduces the potential for expert evaluation to become part of the verification and validation process.

3 A BLUEPRINT FOR END-TO-END PARTICIPATORY DESIGN

We identify the following requirements to extend PD into an *end-to-end* methodology that include the co-design of the automated behaviour of the robot. Such a method needs to allow for the following:

1) systematic observation and study of the use case environment in which the robot is to ultimately be deployed;
2) inclusion and empowerment of relevant stakeholders (users and domain experts) from the initial design phases, such that the design and application of the robot/interaction scenario is co-produced by researchers and stakeholders together;
3) (safe and responsible) evaluation of prototypes in the real-world environment(s) into which the robot is eventually intended to be deployed;
4) inclusion of relevant stakeholders in creation of autonomous robot behaviours, which should utilise interaction data collected in the real world;
5) two-way interaction between the domain expert "teacher" (e.g., a therapist) or designer and robot "learner" such that the teacher can better understand the state of the robot/to what extent learning "progress" is being made and hence adapt their teaching appropriately/flag any significant design issues; and
6) inclusion of relevant stakeholders (e.g., parents of a child in therapy) in (safe) evaluation of autonomous robot behaviours, as they perform in the real world.

Requirements 1 and 2 can be addressed by the typical PD methods discussed in Section 2.1.1, and requirements 3 and 6 can be addressed by carefully designed "in-the-wild" studies. In our work, we therefore look to specifically tackle requirements 4 and 5

TABLE 1 | Key outcomes of and appropriate tools for each stage of LEADOR.

	Outcomes	Tools
1. Problem Definition	Domain understanding	Ethnographic studies, focus groups, brainstorming
2. Interaction Design	Interaction scenario, robot selection/design	Workshop, role-playing, low-tech prototyping
3. System Specification	State-action space for the robot, teaching tools	Brainstorming, behaviour prototyping
4. Technical Implementation	Robot system (sensors and actions), teaching system (authoring tools or learning algorithm)	Software development, laboratory studies, testing workshops
5. Real-World Deployment	Delivering on the application target, autonomous robot	*In situ* teaching by expert

TABLE 2 | Overview of activities undertaken in the two case studies as exemplars for applying our generalised methodology. See Table 4 for a pictorial "storyboard" of this process and the co-design activities undertaken for development of the robot fitness coach.

	School-based educational robot	Gym-based robot fitness coach
Step 1	Decision by researchers based on experience to focus on learning food chain around an educational game for children of age 8–10	Researchers identified the NHS C25K exercise programme based on research goals (longitudinal, real-world HRI) but worked with a fitness instructor to observe typical environment and refine problem definition
Step 2	Decision by researchers to focus on robot-user interaction, with expert only providing robot commands and oversight of the robot behavior to ensure that each action is validated by them. Goal is to evaluate the creation of an autonomous robot	Decision in conjunction with the fitness instructor that the robot would lead exercise sessions (in which he would minimise interaction with exercisers) but that he would provide additional support (e.g., health advice, and stretching) outside of these. Goal is to create and demonstrate an effective, real-world SAR-based intervention *via* PD (as responsible robotics)
Step 3	Using SPARC (Senft et al., 2015) as interaction framework, robot state, and action spaces defined by researchers. Teaching through a GUI on a tablet	Also using SPARC (Senft et al., 2015) the robot state and action spaces as well as the teaching GUI were all co-designed with the fitness instructor
Step 4	Implementation of all the actions and learning algorithm. Prototype evaluation in laboratory. Initial pilot study in schools for evaluating the game with the target population and used as teacher training	Implementation of all the actions and learning algorithm. Fitness instructor provided all dialogues for robot actions. Prototype evaluation was undertaken in the laboratory, and in the final study, gym environment, final robot placement, and system installation details were also decided in conjunction with the fitness instructor
Step 5	Deployment in two local schools with more than 100 children over multiple months. Between-subject evaluation with three conditions: a passive robot, a supervised robot (during the teaching interaction) and an autonomous unsupervised robot	Deployed in to the university gym for teaching and autonomous evaluation through delivery of the C25K programme (27 sessions over 9–12 weeks) to 10 participants. Ran a total of 232 exercise sessions of which 151 were used for teaching the IML system, 32 were used for evaluating the IML system when allowed to run autonomously and 49 were used to test a heuristic-based "control" condition (all testing was within-subject)

by demonstrating how robot automation can be approached as an *in situ*, triadic interaction between domain expert teacher(s), robot learner, and target end user(s). With LEADOR, we showcase how this approach can be integrated into one continuous, end-to-end PD process that satisfies all of the above requirements.

Table 1 summarises the key outcomes of and some potential tools for each stage of LEADOR. **Figure 1** shows how these steps compare to typical PD, as well as who (domain experts and/or engineers) are involved at each stage. Each stage is detailed in full below. **Table 2** shows how these steps have been derived from/were represented in our two foundational studies.

3.1 Step 1: Problem Definition

As noted in **Figure 1**, Step 1 of our method aligns to best practice use of PD as previously demonstrated in social robotics. The purpose of this stage is to generate a thorough understanding of the use context in which the robot is to be deployed and to invite stakeholders to influence and shape the proposed application. It would likely include observations, focus groups, and/or interviews with a variety of stakeholders.

The focus group methodology presented in Winkle et al. (2019a) is one appropriate method that could be used for engaging with stakeholders at this stage because it facilitates expert establishment of non-roboticists, broad discussion of the application context (without presentation of a pre-defined research agenda), participant reflection on the context of use "as is", and researcher-led sharing of technical expertise, followed by detailed consideration and refinement of the research agenda based on researchers and participants now being equal co-designers.

3.2 Step 2: Interaction Design

Similarly to Step 1, Step 2 of our method also aligns to best practice use of PD as previously demonstrated in social robotics. The purpose of this stage is to define and refine the interaction scenario(s) that the proposed robot will engage in and hence the functionalities/capabilities that it might require. The robot platform should also be chosen at this stage. For simplicity, here, we have equated robot platform *choice* with robot platform *design*. Much current social robotics research utilises off-the-shelf robot platforms (e.g., Pepper and NAO from Softbank Robotics), but others focus on the design of new and/or application-specific platforms. Either can be appropriate for LEADOR as long as the choice/design is participatory with stakeholders (for a good example of PD in design of a novel robot, see the work of Alves-Oliveira et al. (2017) on designing the YOLO robot).

Focusing then on more specific application of the robot and the interaction(s) that it should engage in, the methods for PD might include focus groups similarly as those in Step 1 but could also include more novel and/or creative PD activities such as script writing (Björling and Rose, 2019), role playing (including also stakeholder teleoperation of the robot) (Björling and Rose, 2019; Alves-Oliveira et al., 2017), and accessible, "low-tech" prototyping (Valencia et al., 2021).

Note that there is an important interaction design decision to be made here regarding what final deployment of the robot "looks like" in terms of long-term oversight by/presence of domain expert(s) (those involved in its co-design or otherwise) and the role those experts play with regard to the target user. This can be reflected in the teaching interaction setup, specifically with regard to the amount of interaction between the domain expert(s) and target users (see **Figure 2**). For example, it was decided early on in the design of fitness coach robot by Winkle et al. (2020) that there was no intention to ever fully remove the expert presence from the interaction environment. As an alternative, in the work of Senft et al. (2019), the intention from the onset was to create a fully autonomous and independent robot that interacted alone with the target users. Such decisions regarding the role of domain experts would ultimately emerge (explicitly or implicitly) in conjunction with deciding the functionalities of the robot and the further system specification undertaken in Step 3. However, this long-term desired role of the domain expert(s) should be made clear, explicitly, at this stage, such that it can be reflected in the approach to program definition.

3.3 Step 3: System Specification

As shown in **Figure 1**, it is at this stage that our method begins to diverge from the typical PD process, although we continue to utilise PD methods. This step is concerned with co-design of system specifities required to 1) deliver the interaction design resulting from Step 2 and 2) facilitate expert-led teaching phase on real-world deployment that is fundamental to our method (see Step 5). In summary, the aim of this step is to co-design the action space and input space of the robot and the tool(s) that are required to facilitate the bi-directional teaching interaction between the domain expert and the robot. There is also some similarity here to the design process for a WoZ or teleoperated system, which would also require design of the action space of the robot and an interface for (non-roboticist) teleoperation of the robot. The key difference here is the additional requirement to specify the input space of the robot and the choice of teaching tools for the move towards autonomy during Step 5.

3.4 Step 4: Technical Implementation

The main development effort for our method lies in producing the full architecture and tools to allow domain experts to specify autonomous robot behaviour. We note here that the technical implementation required is likely to be greater than that required for a typical WoZ setup and might not be simpler than heuristics-based robot controller.

Four main components need to be developed during this phase:

1) Set of high-level actions for the robot;
2) Set of sensory inputs that will be used to drive the future robot behaviour;
3) A representation of the program which will encode autonomous behaviour; and
4) Expert tools to specify the mapping between the sensory state and the actions.

With our method, the program representation could take the shape of a machine learning algorithm taking inputs from the expert *via* the interface and learning a mapping between the current state of the world when the action was selected and the action itself (the approach taken in our foundational works). Alternatively, the representation could allow the expert to encode a program explicitly, for example, through state machines or trigger-action programming, whilst allowing the expert to update the program in real time and to control the robot actions to ensure that they are constantly appropriate.

A typical automation system would replace the expert tools with an actual definition of the behaviour making use of the program representation to map sensors to actions and define fully an autonomous behaviour. On the other end of the spectrum, a WoZ setup might not need a representation of the program but instead would rely on the interface to display relevant sensory inputs to the wizard (if any) and allow them to select what action to do.

3.5 Step 5: Real-World Deployment and Teaching Phase

Undertaking robot automation (and evaluation) in-the-wild is a key part of LEADOR. To satisfy requirements 5 and 6 as laid out in the introduction, support a mutual shaping approach to robot design, and ensure appropriate robot behaviour, the *teaching phase* should adhere to the following:

1) It must be undertaken *in situ*, i.e., in the context of the final context of use, and with the real target population.
2) It must utilise a domain expert teaching the robot as it delivers on the application interaction.
3) The expert-robot interaction should be bi-directional, i.e., the expert should be able to define and/or refine the autonomous behaviour policy of the robot, whereas the robot informs the expert about its status.

Requirement 1 ensures that the approach is ecologically valid and that the information used by the expert for the automation are suited to the real challenges and idiosyncrasies of the desired context of use.

Requirement 2 ensures that people with domain knowledge can encode that knowledge in the robot. Furthermore, the presence of the expert should be used to ensure that the robot is expressing an appropriate behaviour at all times. As the teaching happens in the real world, with the real users, there is limited space for trial and error. The expert can be used as a safeguard to ensure appropriate robot behaviour even in the initial phases of the teaching.

Requirement 3 ensures that the expert can create a mental model of the robot behaviour. This point is a key difference to non-interactive teaching methods such as the ones based on offline learning (e.g., Sequeira et al., 2016). With the feedback of the robot on its policy (through suggestions or visual behaviour representation), the expert can assess the (evolving) capabilities of the robot and decide what inputs would further improve the policy of the robot.

Finally, during this real-world deployment, if the robot is ultimately expected to interact autonomously/unsupervised, then the expert can use their mental model of the robot behaviour to decide when enough teaching has been done and when the robot is ready to interact autonomously. By relying on online teaching, this decision does not have to be final because the expert could seamlessly step back into the teacher position when the robot interacts with sensitive populations or if the robot requires additional refinement of its policy.

4 FOUNDATIONAL STUDIES

The LEADOR method is primarily derived from two foundational studies made by the authors, which were themselves informed by the previous experiences of authors working with domain experts in the design of social robots. The first one, presented in Senft et al. (2019), explores a study with 75 children on how the teaching interaction could be used to create an autonomous robot behaviour. As shown in **Table 2**, this study did not employ PD, the authors (researchers in HRI) did the early steps by themselves based on their previous related experiences. The second one, presented in Winkle et al. (2020), built on the first study by utilising the same teaching approach to robot automation but incorporating that into an end-to-end PD process to support mutual shaping. The end goal of each study was also slightly different, as Senft et al. (2019) aimed to produce a robot that would ultimately interact with users with little to no further expert involvement. Winkle et al. (2020) also aimed to produce an autonomous robot that would primarily interact 1:1 with users, but with no desire to remove the expert, who would have their own interactions with the users, and/or provide additional teaching to the robot should they deem it necessary.

4.1 Study 1: Evaluating the Teaching Interaction

The goal of this first study was to evaluate if the teaching interaction could be used to create autonomous social behaviours (Senft et al., 2019). This study was designed by the authors, who had experience designing robots for the application domain but did not involve external stakeholders such as teachers.

During the problem definition phase, researchers decided to contextualize the work in robot tutoring for children and explore questions such as how robots can provide appropriate comments to children (both in term of context and time) to stimulate learning. This work was based on experience and knowledge from the researchers about educational robotics.

During the interaction design phase, researchers decided to focus the application interaction around an educational game where children could move animals on a screen and understand food nets. This part included an initial prototype of the game. As the goal was to explore how autonomous behaviours could be created, the teacher was not involved in the game activity, and only the robot was interacting with the child. The robot used was a NAO robot from Softbank Robotics.

In the system specification, the state and action spaces of the interaction were selected. Examples of state include game-related component (e.g., distance between animal) and social dynamics elements (e.g., timing since last action of each agents). The actions of the robot were divided into five categories: encouragements, congratulations, hints, drawing attention, and reminding rules. The teacher-robot interaction used SPARC (Senft et al., 2015).

In the technical implementation phase, the learning algorithm was developed, tested, and interfaced with the other elements of the system. The teaching interface was also created in such a way as to allow the teacher to select actions for the robot to execute and receive suggestions from the robot. At this stage, initial prototypes were tested in laboratory studies and schools.

In the real-world deployment, authors evaluated the system in two different schools with 75 children. The study adopted a between-participant design and explored three conditions: a passive robot, a supervised robot (referring to the teaching interaction), and an autonomous robot (where the teacher was removed from the interaction and the learning algorithm disabled).

Results from the study showed that the teaching interaction allowed the teacher to provide demonstrations to the robot to support learning in the real world. The teacher used the teaching interaction to create a mental model of the robot behaviour. When deployed to interact autonomously, the robot enacted a policy presenting similarities with the one used by the teacher in the teaching phase: the frequency of actions was similar and the robot captured relation and timing between specific events and actions (e.g., a congratulation action should normally be executed around 2 s after an eating event from the actions of a child). Overall, this study demonstrated that human can teach robot social policy from *in situ* guidance.

4.2 Study 2: Teaching Interaction as Participatory Design

The goal of this study was to use the teaching interaction approach to facilitate creation of a fully expert-informed/expert-in-the-loop autonomous socially assistive robot-based intervention for the real world. The fundamental activity to be delivered by robot, the NHS C25K programme, was selected by the researchers based on this research goal, but all study implementation details were decided and designed in conjunction with a domain expert (fitness instructor) throughout. Given the end-to-end and constant expert involvement for this study, there was seamless progression and some overlap between the problem definition, interaction design, and system specification phases, as we present them for LEADOR. A number of co-design activities were undertaken (over a total of six sessions totalling approximately 12.5 h), which ultimately covered all of these key phases, sometimes in parallel, allowing for iteration of the overall study design.

Problem definition was achieved by researchers working with the fitness instructor to 1) understand how a programme like C25K would be delivered by a (human) fitness instructor and 2) explore the potential role a social robot might take in supporting such an intervention. This involved the researchers visiting the university gym and undertaking mock exercise sessions with the instructor, and the instructor visiting the robotics laboratory to see demonstrations of the proposed robot platform and a presentation by the researchers on their previous works and project goals. The robot used was a Pepper robot from Softbank Robotics.

For the interaction design, the researchers and fitness instructor agreed that exercise sessions would be led by the robot and primarily represent robot-user interactions, with the fitness instructor supervising from a distance and only interacting to ensure safety (e.g., in the case of over exertion). As this study also aimed to test (within-subject) the appropriateness of resultant autonomous behaviours, it was decided to purposefully leave the details of the role of the fitness instructor somewhat ambiguous to exercising participants. The instructor was not hidden away during the interaction, and it was clear he was supervising the overall study, but exercisers were not aware of the extent to which he was or was not engaging in teaching interactions with the robot during sessions. As noted in **Section 3**, deciding on what long-term deployment should "look like" in terms of robot-user-expert interactions is a key design requirement at this stage. For the robot fitness coach, we imagined a "far future" scenario, where one of our robot fitness coaches would be installed next to every treadmill on a gym floor, supervised by one human fitness instructor. That instructor would ensure the physical safety of exercisers and still play a role in their motivation and engagement as human-human interaction is known to do. This type of interaction with one expert, multiple robots, and multiple target users is a common goal in many assistive robot applications where some tasks could be automated, but there is a desire to keep an expert presence to, e.g., maintain important human-human interactions and ensure user safety.

The system specification represented somewhat of a "negotiation" between the researchers and the fitness instructor, as he identified the kind of high level action and inputs he felt the robot ought to have, and the researchers identified how feasible that might be for technical implementation. The state space consisted of static and dynamic features that were designed to capture exerciser engagement, task performance, and motivation/personality, all identified by the fitness instructor as being relevant to his decisions in undertaking fitness instruction himself and hence teaching the robot how best to interact with a particular participant. The action space was divided into two categories: task actions and social supporting actions. The task actions were fundamentally set by the C25K programme (i.e., when to run or walk and for how long at a time). The social supporting actions were then broken down into eight sub-categories covering time reminders, social interaction, performance feedback, praise, checking on the user, robot animation, and two proxemics-related actions (leaning towards/away from the user). Importantly, system specification for this study also included co-designing the GUI that would facilitate the bi-directional teaching interaction (also utilising SPARC, Senft et al., 2015) between the robot and the fitness instructor with the fitness instructor himself.

The technical implementation phase essentially mirrored that of Study 1: the learning algorithm was developed, tested, and interfaced with the other elements of the system. The teaching interface was also finalised based on the co-design activities described previously and similarly allowed the fitness instructor to select actions for the robot and to respond to its suggestions. Initial prototypes of both the robot and the GUI were tested in the laboratory studies and the final gym environment.

In the real-world deployment, researchers evaluated the system in a university gym with 10 participants recruited to undertake the 27-session C25K programme over a maximum of 12 weeks. The study adopted a within-subject design and explored three conditions: a supervised robot (referring to the teaching interaction), an autonomous robot (where the fitness instructor was still in position but allowed all learner-suggested actions to auto-execute), and a heuristic-based autonomous robot; a "control" condition for comparing the "teaching interaction as PD" approach to, representing a "next best" alternative for generating expert-informed autonomous behaviour.

Results from the study again demonstrated the feasibility of SPARC and IML for generating autonomous socially assistive robot behaviour suggested that the expert-robot teaching interaction approach can have a positive impact on robot acceptability (by the domain expert and targets users) and that the teaching approach yields better autonomous behaviour that expert informed heuristics as a "next best" alternative for expert-informed autonomous behaviour creation.

4.3 Evidence of Mutual Shaping

Typical PD facilitates mutual shaping as it allows non-roboticist, domain experts to shape research goals, design guidelines, and evaluate robot prototypes, etc. Here, we reflect on observations of mutual shaping effects in our foundational works, specifically resulting from our teaching approach to robot automation.

During our first study, we observed evidences of mutual shaping and the teacher creating a mental model of the robot. For example, our teacher realised with experience that children tended have issues with some aspect of the game (i.e., what food a dragonfly eats). Consequently, she changed her strategy to provide additional examples and support for this aspect of the game. Similarly, the teacher also found that the robot was not initiating some actions often and consequently used these actions more frequently towards the end of the teaching phase to ensure that the robot would exhibit enough of these actions. This exactly evidences the notion that human teachers can tailor their teaching to the progress of a (robot) learner (Bloom, 1984; Thomaz and Breazeal, 2008).

In the second study, we were able to demonstrate mutual shaping in the way the fitness instructor used the robot differently for different participants and/or at different stages of the C25K programme. The longitudinal nature of this study, combined with our approach in supplementing the diadic robot-user interactions with expert-user interactions, meant that the fitness instructor got to know exercise styles/needs of each user and could tailor the behaviour of the robot accordingly. This resulted in the autonomous robot similarly producing behaviour that varied across participants. Similarly, as the programme progressed, the fitness instructor could tailor the behaviour of the robot to reflect the changing exercise demands (e.g., using fewer actions when the periods of running were longer). The flexibility of our approach was also demonstrated when, in response to this increase in intensity, the fitness instructor requested that we add a robot-led cool-down period to the end of each exercise session. This was relatively simple to implement from a technical perspective (an additional "walk" instruction at the end of each session plan) but represented a new part of the session for which there existed no previous training data. As we made this change within the teaching phase (before the switch to autonomous operation), the instructor was able to address this, such that the robot was able to successfully and appropriately support this new cool-down phase when running autonomously.

We also saw an interesting, emergent synergy in the way that the fitness instructor utilised and worked alongside the robot coach. Towards the end of the study, as exercise sessions became more demanding, the fitness instructor took more time at the end of each session to undertake stretching exercises with each participant. This leads to small amounts of overlap between each participant, at which point the fitness instructor would start the next participant warming up with the robot, whilst he finished stretching with the previous participant. We find this to be compelling evidence of the way domain experts will change their practice and/or the way they utilise technological tools deployed into their workplace, particularly when they can be confident in their expectations of how that technology will perform, as is particularly fostered by our approach.

4.4 Interactive Machine Learning for the Teaching Interaction: Opportunities and Limitations

As noted previously, both of our foundational studies utilised IML *via* the SPARC paradigm to facilitate the teaching interaction. From a technical perspective, our foundational studies demonstrate the feasibility and relative effectiveness (in terms of teaching time) of this approach. Fundamentally, LEADOR is agnostic with regard to the specific computational approach to facilitating the teaching interaction, but we find IML to be a particularly compelling solution, in line with the overall aims of the method, as it makes for an intuitive bi-directional teaching interaction for the domain expert. Specifically, through one single interface, they can see what the robot intends to do (and potentially why) before that action is executed, improving their understanding of the learning progress of the robot, and instantiate teaching exemplars in real time, informed by that understanding as well as the instantaneous requirements of the application task.

However, here, we draw attention to one key limitation regarding expert-robot interactions and assessment when using IML. An important element of mutual shaping not considered here is if/how/to what extent the suggestions made by the learning robots may have influenced the domain experts. For example, had the learning robots not been making suggestions, such that the robot was entirely controlled/teleoperated by the experts, would the action distribution and timing of actions remained the same? Further, if the experts did not have the ability to actively reject suggestions (indicating that the learner was not producing appropriate robot behaviour), then would they still have *post hoc* identified those actions as being inappropriate?

This is particularly interesting given the high number of suggested actions still being rejected at the end of the training phase, in both of our foundational studies, immediately followed by seemingly appropriate robot behaviour that was positively evaluated by the experts themselves during autonomous operation. Success of our approach inherently assumes that the domain expert/system "teacher" would provide a "correct" and fairly consistent response; i.e., that they 1) can correctly assess the quality of each action suggested by the robot and make an informed about whether this action should be executed and 2) are always able to ensure that required robot actions are executed in a timely fashion. With SPARC, these robot suggestions are the main means to help the expert create a mental model of the robot behaviour. Consequently, whilst our results demonstrate that the IML does fundamentally "work" for automating robot behaviour and that our domain experts did construct a mental model of the behaviour of the robots, there remains an open question regarding how the robot could improve the transparency of its behaviour to actively support mental model creation for the teacher.

5 DISCUSSION

5.1 A Flexible and Effective Method for Automating Social Robots

We suggest that LEADOR can be used to design robots for a variety of interaction settings, in terms of the required autonomy and the nature of expert-robot-user interactions long-term. We propose two axes to describe the different types of interaction that might be desired, based on the application (**Figure 3**). A first axis describes the extent to which the domain expert(s) and user(s) are expected to

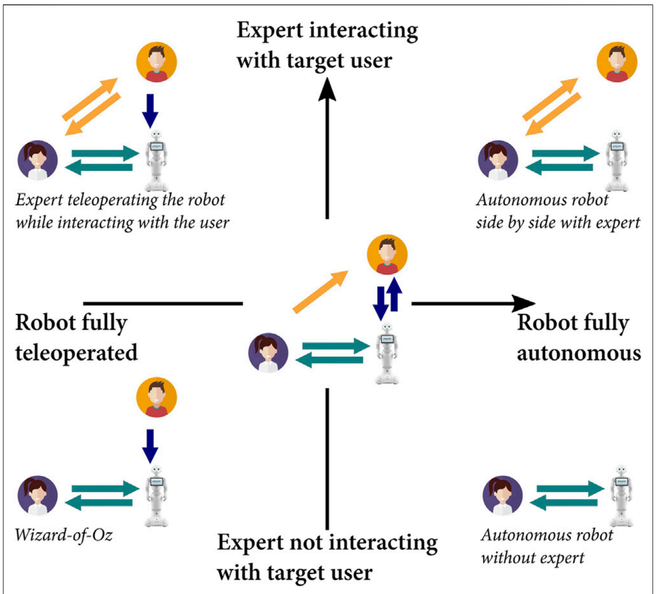


FIGURE 3 | Two-dimensional representation for visualising the different types of long-term expert-robot-user interactions that a social robot might be designed for, all of which LEADOR can facilitate. Note that this is not a discrete space, and LEADOR specifically makes it possible to move along these axes upon real-world deployment. People vector created by studiogstock - www.freepik.com.

interact long term, as a supplement to the robot-user interaction(s). The second axis reflects the autonomy of the robot, from full supervision (teleoperation) to full autonomy. These two axes are independent as, for example, cases exist where the expert might be continuously interacting with the target users, whilst continuing or not to supervise and/or improve the autonomous behaviour of the robot long term. In addition, these axes do not represent a discrete space, as the teaching interaction element of LEADOR specifically makes it possible to move along either axis at any point during real-world deployment.

The robots developed in our foundational studies demonstrate this flexibility and exist in slightly different spaces on these axes. The educational robot by Senft et al. (2019) is an example of an autonomous robot operating without the expert, and the teaching interaction represented a typical Wizard-of-Oz setup, i.e., there was no interaction. The robot fitness coach by Winkle et al. (2020) is closer to an autonomous robot operating side by side with the domain expert, and the teaching interaction utilised some interactions between the expert and the users (although this was undertaken *outside* of direct teleoperation).

The two foundational studies also demonstrate and evaluate different, complimentary elements of the effectiveness of LEADOR for designing social robots. More specifically, Senft et al. (2019) fundamentally demonstrated the practical feasibility of the teaching interaction for creating appropriate autonomous behaviour. After a teaching phase with 25 children, the robot was deployed autonomously and without expert supervision. It displayed a similar policy to when it was supervised, for example, capturing connections between some events and actions with appropriate timing. However, it was not using a PD approach from the onset, if LEADOR had been applied, then teachers would have been involved more thoroughly in the game design and the interface development.

Whilst Winkle et al. (2020) again demonstrated similarity between supervised and autonomous behaviour, this work also specifically demonstrated that the teaching interaction resulted in a better autonomous robot than an expert-informed heuristic based alternative. In addition, the work specifically explored to what extent the overall LEADOR could support mutual shaping and influence robot acceptability. To this end, as shown in **Figure 4**, the significant co-design work undertaken by the domain expert seems likely to have contributed to the high level of ownership he seemed to feel towards the system, and the way in which he conceptualised the robot, throughout, as an independent agentic colleague he was training. When asked whether he perceived Pepper as more of a tool or a colleague, the fitness instructor commented *"It was definitely more of a colleague than a tool [. . .] I like to think her maybe early bugs or quirks definitely gave her a bit more of a personality that maybe I held on to"*. In addition, when evaluating the performance of the robot, the instructor also reflected on the difference between how the robot might behave in comparison to himself: *"Pepper's suggestions might not be what *I* would say in that exact same situation; however, it does not mean that what was said or suggested was wrong"*. This gives credibility to the suggestion that LEADOR can be used to create robots that do not simply attempt to imitate or replicate the domain expert directly but instead play a distinct but complimentary role alongside that domain expert in delivering an assistive intervention.

The feedback of the fitness instructor also suggested that the use of the robot did not prevent him from still developing a working relationship with the exercisers or from having a positive impact on their motivation, as he *"did care about their progress and their health"*. This appears to be true on the side of the exerciser, too, because their evaluations suggested they perceived the fitness instructor and the robot as playing distinct but complimentary roles in their undertaking of and engagement with the prescribed exercise programme: *"Pepper was a good instructor and positively motivated my runs. The role of Don [the fitness instructor] assisted this in that having him there meant I could follow the robot's instructions safe in the knowledge that there was some support there should anything go wrong!"*

In summary, the fitness coach robot example therefore demonstrates the end-to-end PD element of LEADOR, how this seemingly contributes to robot acceptability by both domain experts and target users, and can successfully facilitate meaningful triadic (domain expert-robot-user) interactions in human-centred domains where there might be a desire to reduce domain expert workload without ever removing them from the interaction completely. As such, Winkle et al. (2020) might be seen as a first attempt to fully implement LEADOR ahead of refinement for presentation as a generalisable methodology.

5.2 Supporting "Responsible by Design" Robotics

The Foundation for Responsible Robotics (FRR) defines responsible robotics as "the responsible design, development, use,

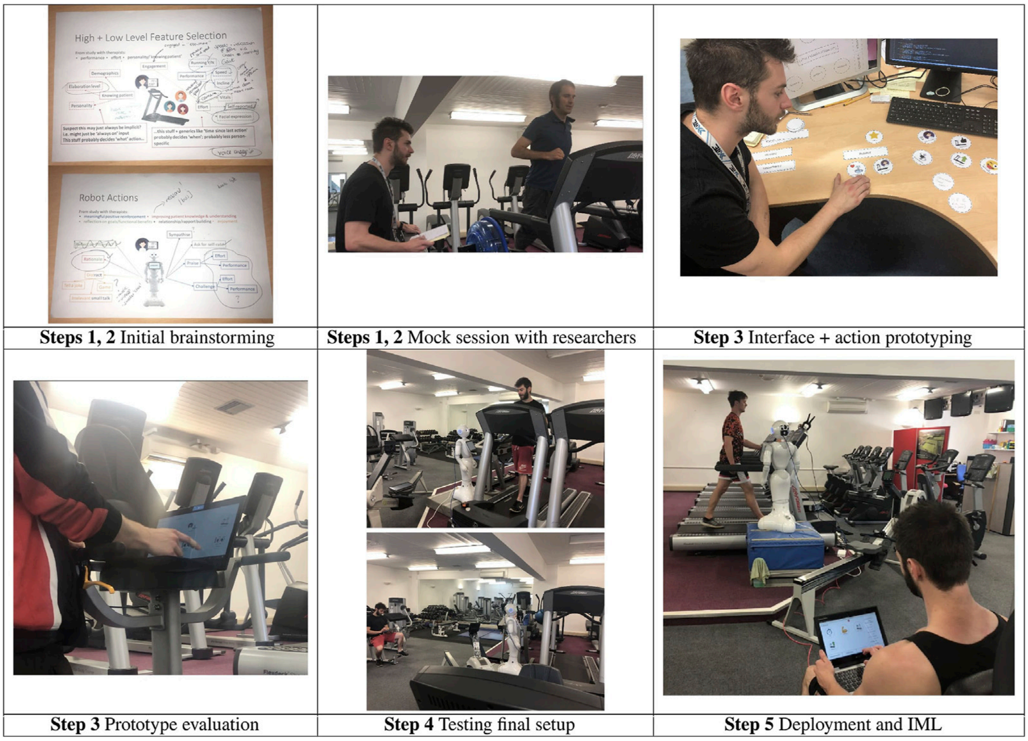


FIGURE 4 | Pictorial representations of the participatory design activities and final teaching setup undertaken in application of our method to the robot fitness coach by Winkle et al. (2020), as per **Table 2** with reference to Steps 1–5 of our method as per **Figure 1**.

implementation, and regulation of robotics in society"[1]. Concerning research and development, the FRR demonstrates a significant overlap with the goals of mutual shaping and, hence, our goals in proposing LEADOR: "Responsible robotics starts before the robot has been constructed. Ethical decision-making begins in the R&D phase. This includes the kind of research practices that are employed, ensuring that a diverse set of viewpoints are represented in the development of the technology, using methods of development and production that are sustainable, and taking into consideration the impact that the technology will have on all stakeholders to mitigate harm preemptively rather than after the fact."

A significant number of attempts to more formally define the ethical design and development have taken the form of published principles of AI and robotics [2], many of which similarly identify the importance of engaging (non-roboticist) users and domain experts in robot design and evaluation processes. Arguably, one of the more practical resources is the British standard BS8611-2016 *Guide to the Ethical Design and Application of Robots and Robotic Systems* (BSI, 2016), which explicitly identifies ethical risks posed by robots, mitigating strategies and suggested methods for verification and validation. Notably, the standard suggests that a number of the identified ethical hazards might be verified and validated through *expert guidance* and *user validation*. Through LEADOR, such guidance and validation is inherently "built-in" to the design and development process. On the basis of this, we posit that, in supporting a mutual shaping approach to robot development, and specifically by "opening up" robot automation to non-roboticists (such that they can contribute to robot design and automation but also better understand robot capabilities and limitations), LEADOR also represents a concrete implementation of a *responsible robotics* approach and offers a practical way to create social robots with expert guidance and user validation being inherent to the development process. Consequently, whilst the program is evaluated by its designers, these designers are the domain experts and thus the best persons to assess whether the robot behaviour is successful or not.

5.3 Future Development

5.3.1 Inclusion of Application Targets in Design, Automation, and Evaluation

A key limitation in both of our foundational works was the lack of including target users during the design processes. This is partly because both of these works are concerned in the development of robots that would be assisting the domain expert practitioners (i.e., a teaching assistant and a fitness instructor), and so, it made sense to focus on working with such experts as co-designers of the system. However, as discussed in the introduction, inclusion of all stakeholders is a key aim of mutual shaping approaches to robot design/development.

A desire to include target users in the design and evaluation of the robot would raise the interesting question of how target users, who

[1]https://responsiblerobotics.org/

[2]http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html

are expecting to beneficiaries of the interaction, could design the robot. In a number of situations where the robot is expected to provide support or additional knowledge, including target users in the co-design of the action state, for example, could be either complex or negate the *perception* of the robot as an agent. This also overlaps with discussions in contemporary PD works regarding the "legitimacy" of different PD participants and specifically with the idea of participants *learning* from the researchers as a pre-requisite for becoming such (Ehn, 2008). In the first instance, however, as an obvious extension to LEADOR, target users could certainly be included in preliminary testing of those actions designed with a domain expert.

5.3.2 Alternative Teaching/Learning Interactions

The method presented in this paper focused on a teaching phase where a domain expert teaches the robot how to interact with a target user, with the target user unaware of the extent to which the expert is involved in the robot behaviour. It abstracts away the type of learning used as each situation has different constraints and requires variations on the teaching interaction and learning algorithm. Consequently, LEADOR might not be applicable directly to every situation. Future work should explore the applicability of LEADOR with other interaction designs not explored in our foundational studies and explore combination with other methodologies to extend this applicability whilst maintaining the key tenets regarding expert involvement and *in situ* robot design, teaching, and testing.

Whilst situations such as therapy or education require the expert and target user to be different persons, a large number of other domains relax this constraint. For example, an elderly at home could have a robot carer and teach the robot how to support them in their daily activity. In this case, the target user is the person knowing best their needs and as such would be the perfect expert. LEADOR would be highly applicable to this situation as the target users could be involved early in the design process, help specify the state and action spaces and tools that they would need, and finally teach *in situ* their robot how to interact whilst benefiting from the interaction themselves.

Alternatively, building on the previously noted limitation regarding target user inclusion, applications where the robot is to play more of a *peer* role, rather than an expert authority might be best achieved by having one target user teach the robot how to interact with another target user. This might be particularly appropriate for, e.g., allowing teenagers to automate companion robots that support the mental health of teenagers (Björling and Rose, 2019). This raises a number of interesting research questions regarding how the teaching interaction might impact on the teacher's (self-)understanding of the application domain, representing another aspect of mutual shaping that could be considered in more detail in future works.

An alternative, exciting teaching interaction is having the teaching phase being open and transparent to the target user. Teaching robots could be similar to how adult teach children to interact, by providing explicit feedback guideline openly in the social environment. This situation raises a number of open questions such as to what extent having the expert providing feedback to the robot could impact the ascribed agency of the robot or how could the target user be included in telling the robot how best to help them. We have good evidence from our work (Winkle et al., 2020) that such open interaction would not "break the illusion" of the robot being an independent (credible) social agent. Further, previous work suggests that robot users value the human developers "behind" the robot, because it is their "genuine intentions" that underlie the social and assistive behaviours of robot (Winkle et al., 2019b). In sensitive application environments such as the previously mentioned teenage mental health support, such openness may indeed be crucial to robot effectiveness and acceptability (Björling et al., 2020).

However, these alternative teacher/learner configurations need to account for the existing practical constraints of using reinforcement learning (RL) in human-robot interaction. Indeed, in the context of HRI, RL faces two main issues: 1) the large number of data points required to improve the policy (which have to come from real-world interaction) and 2) the risks posed by the RL "exploration" in the real HRI, where the RL algorithm might suggest actions that are inappropriate in a given context.

In our two studies, the domain expert also acted as a "gate keeper" for the suggestions of the robot and as a general safety net, able to intervene if the autonomous robot behaviour was inappropriate. Likewise, when applying LEADOR in other scenarios, adequate safeguarding needs to be in place, until further research on RL can provide adequate safety guarantees. Alternatively, the expert could serve early on to help create an initial safe and effective policy by providing a high amount of guidance. Then, in the second phase, the expert could revert only to the "gate keeper" role, working as a safeguard to ensure that the policy of the robot has a minimum efficacy whilst letting the robot self-improve. Finally, when the robot reaches a sufficient expertise in the interaction, it could be left to fine-tune its policy with less supervision.

6 CONCLUSION

In this article, we present LEADOR, a method for end-to-end PD of autonomous social robots that supports a mutual shaping approach to social robotics. This general method is derived from two independent foundational studies and represents a culmination of the experiences of authors working with domain experts in the development of autonomous socially assistive robots. We describe the activities undertaken in those studies to demonstrate how the method has been derived and give tangible examples of how it might be applied. Together, we suggest that these foundational studies also demonstrate both the feasibility and the value of the approach, because both resulted in acceptable, autonomous, and effective socially assistive robots successfully utilised in complex real-world environments.

The first key contribution of LEADOR is to make robot *automation* participatory, such that non-roboticist, domain experts can contribute directly to generating autonomous robot behaviours. This particularity compliments more typical use of PD, e.g., generating the initial robot design guidelines or evaluation robot prototypes. We achieve this expert-led automation by utilising a *teaching interaction*, whereby the domain expert(s) can directly define and refine the autonomous behaviour of the robot through a teaching interface. Both of our foundational studies utilised IML and the SPARC paradigm (Senft et al., 2015), which we suggest is particularly well suited to the overall method goals; therefore, we particularly reflect on this approach and its benefits, challenges, and limitations.

However, whilst we refer to this as a teaching interaction, because the domain expert is "teaching" the robot how to behave, our method is agnostic as to the specific technical approach taken (e.g., machine learning and authoring) to facilitate it.

The second key contribution of our LEADOR is to facilitate a mutual shaping approach throughout robot development. This is achieved, first, by the increased domain expert participation in robot automation as described above. In addition, however, our integration of the teaching interaction into real-world robot deployment means that this automation of robot behaviour can actually be informed by and reflect the complex and nuanced realities of the real-world context, capturing the tacit and intuitive responses of the expert to real-world social dynamics. Given that teaching can be re-convened at any time, the method also facilitates the updating of robot behaviours in response to these evolving dynamics or new emerging dynamics, i.e., observation of mutual shaping effects. More generally, the *in situ* robot deployment and expert teaching role maximise the opportunity to identify and understand such mutual shaping effects to better evaluate the overall impact and efficacy of the robot for the proposed application.

In facilitating end-to-end PD and mutual shaping, we also suggest that our method inherently supports responsible robotics, by design. Specifically, it allows for a diverse set of viewpoints to be represented in the development of the technology and for preemptive consideration of the impact that technology will have on stakeholders. Finally, on a practical level, we also suggest our method can better facilitate multi-disciplinary working because it systematically combines PD and technical development such that non-roboticist researchers and stakeholders are no longer excluded from any stage of the development process.

In summary, we suggest that LEADOR is an all-around effective approach for creating socially intelligent robots, as *practical* as it is *responsible* in facilitating the creation of expert-informed, intuitive social behaviours. We identify a number of areas for potential future development, which we hope will be of interest to other roboticists in refining the method further and working further towards democratisation of robot design and development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Plymouth and the University of the West of England Research Ethics Committees as appropriate. Written informed consent to participate in these studies were provided by the legal guardian/next of kin of the participants. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

KW and ES led foundational studies 1 and 2 from which this work is derived, both of which were (independently) conducted in close collaboration with SL. KW and ES led on derivation of the generalisable method based on their shared experiences, with all authors contributing to reflections on the foundational studies, resultant implications for the generalisable methodology and producing the final manuscript.

FUNDING

This work was partially supported by the EPSRC via the Centre for Doctoral Training in Future Autonomous and Robotic Systems (FARSCOPE) at the Bristol Robotics Laboratory, University of the West of England and University of Bristol (grant number EP/L015293/1), partially funded by the EU FP7 DREAM project (grant no. 611391) and partially funded by the KTH Digital Futures Research Center.

ACKNOWLEDGMENTS

We wish to acknowledge our PhD supervisors, Paul Bremner, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, Tony Belpaeme, and Paul Baxter with whom we collaborated on the foundational studies and previous experiences that informed this work. In addition, we wish to acknowledge the two domain experts, Madeleine Bartlett and Donald Knight, who took part in our foundational works whose reflections contributed to our refinement of the method.

REFERENCES

Alves-Oliveira, P., Arriaga, P., Paiva, A., and Hoffman, G. (2017). "YOLO, a Robot for Creativity," in Proceedings of the 2017 Conference on Interaction Design and Children (IDC '17), New York, NY, United States (New York, NY, USA: Association for Computing Machinery), 423–429. doi:10.1145/3078072.3084304

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the People: The Role of Humans in Interactive Machine Learning. *AIMag* 35, 105–120. doi:10.1609/aimag.v35i4.2513

Arkin, R. C., Fujita, M., Takagi, T., and Hasegawa, R. (2001). "Ethological Modeling and Architecture for an Entertainment Robot," in Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164), Seoul, Korea, May 2001, 453–458. doi:10.1109/ROBOT.2001.932592

Azenkot, S., Feng, C., and Cakmak, M. (2016). "Enabling Building Service Robots to Guide Blind People a Participatory Design Approach," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, March 2016, 3–10. doi:10.1109/hri.2016.7451727

Beer, J. M., Smarr, C.-A., Chen, T. L., Prakash, A., Mitzner, T. L., Kemp, C. C., and Rogers, W. A. (2012). "The Domesticated Robot," in HRI '12: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, United States, March 2012 (New York, NY, USA: ACM), 335–342. doi:10.1145/2157689.2157806

Björling, E. A., Thomas, K., Rose, E. J., and Cakmak, M. (2020). Exploring Teens as Robot Operators, Users and Witnesses in the Wild. *Front. Robot. AI* 7. doi:10.3389/frobt.2020.00005

Björling, E., and Rose, E. (2019). Participatory Research Principles in Human-Centered Design: Engaging Teens in the Co-design of a Social Robot. *Mti* 3, 8. doi:10.3390/mti3010008

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-To-One Tutoring. *Educ. Res.* 13, 4–16. doi:10.3102/0013189x013006004

Bratteteig, T., and Verne, G. (2018). "Does AI Make PD Obsolete? Exploring Challenges from Artificial Intelligence to Participatory Design," in Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2, Genk, Belgium, August 2018, 1–5.

BSI (2016). [Dataset]. BS 8611 - Robots and Robotic Devices: Guide to the Ethical Design and Application of Robots and Robotic Systems (No. BS 8611:2016). BSI Standards Publication.

Cao, Y., Wang, T., Qian, X., Rao, P. S., Wadhawan, M., Huo, K., et al. (2019). "Ghostar: A Time-Space Editor for Embodied Authoring of Human-Robot Collaborative Task with Augmented Reality," in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, New Orleans, Louisiana, United States, October 2019, 521–534. doi:10.1145/3332165.3347902

Chang, W.-L., and Šabanović, S. (2015). "Interaction Expands Function," in HRI '15: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, Oregon, United States, March 2015 (New York, NY, USA: ACM), 343–350. doi:10.1145/2696454.2696472

de Graaf, M. M. A., Allouch, S. B., and Klamer, T. (2015). Sharing a Life with Harvey: Exploring the Acceptance of and Relationship-Building with a Social Robot. *Comput. Hum. Behav.* 43, 1–14. doi:10.1016/j.chb.2014.10.030

Ehn, P. (2008). "Participation in Design Things," in Proceedings of the Tenth Participatory Design Conference (PDC), Bloomington, Indiana, United States, October 2008 (Bloomington, Indiana, USA: ACM Digital Library), 92–101.

Fails, J. A., and Olsen, D. R., Jr (2003). "Interactive Machine Learning," in Proceedings of the 8th International Conference on Intelligent User Interfaces, Miami, Florida, United States, January 2003, 39–45. doi:10.1145/604045.604056

Fleron, B., Rasmussen, R., Simonsen, J., and Hertzum, M. (2012). "User Participation in Implementation," in Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases-Volume 2, Roskilde, Denmark, August 2012, 61–64. doi:10.1145/2348144.2348164

Forlizzi, J. (2007). "How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the home," in 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Arlington, Virginia, United States, March 2007, 129–136.

Gao, Y., and Huang, C.-M. (2019). "Pati: a Projection-Based Augmented Table-Top Interface for Robot Programming," in Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, California, United States, March 2019, 345–355.

Huang, J., and Cakmak, M. (2017). "Code3: A System for End-To-End Programming of mobile Manipulator Robots for Novices and Experts," in 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Vienna, Austria, March 2017 (IEEE), 453–462.

Jenkins, S., and Draper, H. (2015). Care, Monitoring, and Companionship: Views on Care Robots from Older People and Their Carers. *Int. J. Soc. Robotics* 7, 673–683. doi:10.1007/s12369-015-0322-y

Knox, W. B., Spaulding, S., and Breazeal, C. (2014). "Learning Social Interaction from the Wizard: A Proposal," in Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada, July 2014.

Lee, H. R., Šabanović, S., Chang, W.-L., Nagata, S., Piatt, J., Bennett, C., and Hakken, D. (2017). "Steps toward Participatory Design of Social Robots," in HRI '17:Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, March 2017 (New York, NY, USA: ACM), 244–253. doi:10.1145/2909824.3020237

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., and Alami, R. (2017). Artificial Cognition for Social Human-Robot Interaction: An Implementation. *Artif. Intelligence* 247, 45–69. doi:10.1016/j.artint.2016.07.002

Leonardi, N., Manca, M., Paternò, F., and Santoro, C. (2019). "Trigger-action Programming for Personalising Humanoid Robot Behaviour," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, United Kingdom, May 2019, 1–13. doi:10.1145/3290605.3300675

Liu, P., Glas, D. F., Kanda, T., and Ishiguro, H. (2016). Data-Driven HRI: Learning Social Behaviors by Example from Human-Human Interaction. *IEEE Trans. Robot.* 32, 988–1008. doi:10.1109/tro.2016.2588880

Louie, W.-Y. G., Li, J., Vaquero, T., and Nejat, G. (2014). "A Focus Group Study on the Design Considerations and Impressions of a Socially Assistive Robot for Long-Term Care," in The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, Scotland, United Kingdom, August 2014, 237–242. doi:10.1109/ROMAN.2014.6926259

Louie, W.-Y. G., and Nejat, G. (2020). A Social Robot Learning to Facilitate an Assistive Group-Based Activity from Non-expert Caregivers. *Int. J. Soc. Robotics* 12, 1–18. doi:10.1007/s12369-020-00621-4

Paxton, C., Hundt, A., Jonathan, F., Guerin, K., and Hager, G. D. (2017). "Costar: Instructing Collaborative Robots with Behavior Trees and Vision," in 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, June 2017 (IEEE), 564–571. doi:10.1109/icra.2017.7989070

Porfirio, D., Fisher, E., Sauppé, A., Albarghouthi, A., and Mutlu, B. (2019). "Bodystorming Human-Robot Interactions," in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, New Orleans, Louisiana, United States, October 2019, 479–491. doi:10.1145/3332165.3347957

Porfirio, D., Stegner, L., Cakmak, M., Sauppé, A., Albarghouthi, A., and Mutlu, B. (2021). "Figaro: A Tabletop Authoring Environment for Human-Robot Interaction," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 2021. doi:10.1145/3411764.3446864

Riek, L. (2012). Wizard of Oz Studies in Hri: a Systematic Review and New Reporting Guidelines. *Jhri* 1, 119–136. doi:10.5898/jhri.1.1.riek

Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., et al. (2011). "Child-robot Interaction in the Wild: Advice to the Aspiring Experimenter," in Proceedings of the 13th International Conference on Multimodal Interfaces, Alicante, Spain, November 2011, 335–342.

Šabanović, S. (2010). Robots in Society, Society in Robots. *Int. J. Soc. Robotics* 2, 439–450. doi:10.1007/s12369-010-0066-7

Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M., and Belpaeme, T. (2019). Teaching Robots Social Autonomy from *In Situ* Human Guidance. *Sci. Robot* 4, eaat1186. doi:10.1126/scirobotics.aat1186

Senft, E., Baxter, P., Kennedy, J., and Belpaeme, T. (2015). "Sparc: Supervised Progressively Autonomous Robot Competencies," in International Conference on Social Robotics, Paris, France, October 2015 (Springer), 603–612. doi:10.1007/978-3-319-25554-5_60

Sequeira, P., Alves-Oliveira, P., Ribeiro, T., Di Tullio, E., Petisca, S., Melo, F. S., et al. (2016). "Discovering Social Interaction Strategies for Robots from Restricted-Perception Wizard-Of-Oz Studies," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, March 2016 (IEEE). doi:10.1109/hri.2016.7451752

Simonsen, J., Bærenholdt, J. O., Büscher, M., and Scheuer, J. D. (2010). *Design Research: Synergies from Interdisciplinary Perspectives*. Abingdon-on-Thames, United Kingdom: Routledge.

Simonsen, J., and Hertzum, M. (2012). Sustained Participatory Design: Extending the Iterative Approach. *Des. Issues* 28, 10–21. doi:10.1162/desi_a_00158

Simonsen, J., and Robertson, T. (2012). *Routledge International Handbook of Participatory Design*. Cambridge, MA, United States: MIT Press.

Simonsen, J., Svabo, C., Strandvad, S. M., Samson, K., and Hertzum, M. (2014). *Situated Design Methods*. Abingdon-on-Thames, United Kingdom: Routledge.

Sussenbach, L., Riether, N., Schneider, S., Berger, I., Kummert, F., Lutkebohle, I., and Pitsch, K. (2014). "A Robot as Fitness Companion: Towards an Interactive Action-Based Motivation Model," in The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, Scotland, United Kingdom, August 2014, 286–293. doi:10.1109/ROMAN.2014.6926267

Thomaz, A. L., and Breazeal, C. (2008). Teachable Robots: Understanding Human Teaching Behavior to Build More Effective Robot Learners. *Artif. Intelligence* 172, 716–737. doi:10.1016/j.artint.2007.09.009

Valencia, S., Luria, M., Pavel, A., Bigham, J. P., and Admoni, H. (2021). "Co-designing Socially Assistive Sidekicks for Motion-Based AAC," in HRI '21: Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, March 2021 (New York, NY, USA: Association for Computing Machinery), 24–33. doi:10.1145/3434073.3444646

Winkle, K., Caleb-Solly, P., Turton, A., and Bremner, P. (2019a). Mutual Shaping in the Design of Socially Assistive Robots: A Case Study on Social Robots for Therapy. *Int. J. Soc. Robotics* 12, 847–866. doi:10.1007/s12369-019-00536-9

Winkle, K., Caleb-Solly, P., Turton, A., and Bremner, P. (2018). "Social Robots for Engagement in Rehabilitative Therapies," in HRI '18:Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, United States, March 2018 (New York, NY, USA: ACM), 289–297. doi:10.1145/3171221.3171273

Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A., and Bremner, P. (2019b). "Effective Persuasion Strategies for Socially Assistive Robots," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, South Korea, March 2019, 277–285. doi:10.1109/HRI.2019.8673313

Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A., and Bremner, P. (2020). "*In-situ* Learning from a Domain Expert for Real World Socially Assistive Robot Deployment," in Proceedings of Robotics: Science and Systems 2020. doi:10.15607/rss.2020.xvi.059

Wu, Y.-H., Fassert, C., and Rigaud, A.-S. (2012). Designing Robots for the Elderly: Appearance Issue and beyond. *Arch. Gerontol. Geriatr.* 54, 121–126. doi:10.1016/j.archger.2011.02.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/frobt.2021.719154

Aerial Flight Paths for Communication

Alisha Bevins and Brittany A. Duncan**

NIMBUS Lab, Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE, United States

This article presents an understanding of naive users' perception of the communicative nature of unmanned aerial vehicle (UAV) motions refined through an iterative series of studies. This includes both what people believe the UAV is trying to communicate, and how they expect to respond through physical action or emotional response. Previous work in this area prioritized gestures from participants to the vehicle or augmenting the vehicle with additional communication modalities, rather than communicating without clear definitions of the states attempting to be conveyed. In an attempt to elicit more concrete states and better understand specific motion perception, this work includes multiple iterations of state creation, flight path refinement, and label assignment. The lessons learned in this work will be applicable broadly to those interested in defining flight paths, and within the human-robot interaction community as a whole, as it provides a base for those seeking to communicate using non-anthropomorphic robots. We found that the Negative Attitudes towards Robots Scale (NARS) can be an indicator of how a person is likely to react to a UAV, the emotional content they are likely to perceive from a message being conveyed, and it is an indicator for the personality characteristics they are likely to project upon the UAV. We also see that people commonly associate motions from other non-verbal communication situations onto UAVs. Flight specific recommendations are to use a dynamic retreating motion from a person to encourage following, use a perpendicular motion to their field of view for blocking, simple descending motion for landing, and to use either no motion or large altitude changes to encourage watching. Overall, this research explores the communication from the UAV to the bystander through its motion, to see how people respond physically and emotionally.

Keywords: drone, gesture, human-robot interaction, communication, small UAS

OPEN ACCESS

Edited by:
Ginevra Castellano,
Uppsala University, Sweden

Reviewed by:
Autumn Edwards,
Western Michigan University,
United States
Ilaria Torre,
Royal Institute of Technology, Sweden

****Correspondence:***
Alisha Bevins
abevins@cse.unl.edu
Brittany A. Duncan
bduncan@unl.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *01 June 2021*
Accepted: *16 November 2021*
Published: *07 December 2021*

Citation:
Bevins A and Duncan BA (2021) Aerial Flight Paths for Communication. Front. Robot. AI 8:719154. doi: 10.3389/frobt.2021.719154

1 INTRODUCTION

As UAVs increase in popularity and functionality, they are becoming easier to obtain and significantly more visible in standard occurrences for the general public. In addition to the increase in visibility to the public in everyday occurrences, they are being used in many professional environments such as disaster relief, agriculture, and product delivery. One of the problems with increased visibility and use is that not everyone who comes in contact with the UAV will have context for its purpose or current task. This becomes an even larger issue when a malfunction or abnormality occurs. UAV manufacturers, programmers, and users need to be able to understand how they can expect the uninformed person to react to their vehicle. In addition to this, a bystander needs to be able to understand what is occurring to minimize concern and unnecessary intervention.

The main purpose of this work is to inform future researchers, and UAV developers, about how participants perceive UAV paths. This includes what they believe the system to be communicating, the most important components of the flight paths, and their intended reactions based on those communications. To address these issues, we first explore how consistently people label motions (Phase 0). Those labels were then presented to participants to create their own motions to see if there were inherent similarities in these motions (Phase 0). A combined set of motions were then presented to new participants to see if user generated paths had increased label agreement (Phase 1–3). Finally, states which were more effective at generating responses were presented to a final set of participants to understand whether their created motions would align with the expected path characteristics from earlier phases (Phase 4). **Figure 1** further introduces the phases and how they will be presented in this paper, in addition to showing prior contributing works.

Overall lessons from this work indicate that:

- frequent motions or gestures applied in non-UAV situations are associated and understood on UAVs,
- landing is conveyed by direct movements with an altitude change,
- people will follow a UAV's path when the motion approaches and then retreats towards a location when in the absence of altitude changes, and
- flights across an area are likely to cause participants to avoid the vehicle or that area (regardless of the altitude).

We found that simpler motions are more likely to have consistent interpretation across participants. Considering the most basic flight paths, people took the front-back motion on the y-axis to mean to follow the vehicle, a side to side motion focused on the x-axis to stay back (or to not follow it), and an up-down motion on the z-axis to mean landing. We also found that NARS can be an indicator of how people expect to react, if they are likely to expect a negative message to be conveyed, and their expectation for the UAV to have negative personality traits.

2 LITERATURE REVIEW

When considering the topics discussed in this paper, the related work is broad and inherits best practices across many fields. This chapter discusses the most relevant work when developing the studies and provides context to those hoping to adopt these practices in the future.

2.1 Social UAVs

The work of social UAVs, which we will define as "UAVs that will operate in spaces used by and necessitate communication with human bystanders," has been expanding rapidly in recent years. This has lead to (Funk, 2018) providing a comprehensive overview of UAVs as flying interfaces, and (Baytas et al., 2019) providing design recommendations for UAVs in inhabited environments. A significant finding from (Baytas et al., 2019) discusses the idea of providing future work on "Intuitive Comprehension" of UAV movements, which means understanding what a UAV is trying to convey without additional explanation. A more comprehensive discussion of social uses for UAV systems can be found in these works.

2.2 UAV Communication

(Cauchard et al., 2015; Obaid et al., 2016) have examined different methods to facilitate communication from the human to the UAV. In the work presented here we are more interested in what a UAV can communicate to a person who may or may not be its operator. This can be achieved through a variety of methods, with the most popular discussed further here.

2.2.1 Video, Lights, and Stereo

Audio or video methods can be very direct in their communication by providing speech, either verbal or written, or figures. Attaching a projector onto a UAV is a common video communication method, as demonstrated by (Matrosov et al., 2016; Nozaki, 2014; Scheible et al., 2013). These projects typically project text or video onto an arbitrary object and can also include an interface to allow user control of the display (Matrosov et al., 2016). Merged these uses by creating interfaces projected onto the ground that allowed interactions using motions of a foot. Another visual modality demonstrated for UAV communications are lights (Szafir et al., 2015). Showed the ability to convey robot flight intentions at a glance, specifically to better express directionality. They found participants were able to better distinguish robot predictability over baseline flight behaviors when given four different signal designs.

On the audio side, providing speech to observers is as straightforward as attaching a speaker (Yamazaki et al., 2019). Demonstrated one strong use case by attaching a speaker and microphone system on a UAV that would make sounds for natural disaster victims to react to, and then capture their vocal reactions.

Although most of these studies were more qualitative in nature and had limited participants included, they do show the capability of direct communication from a UAV to a bystander. Unfortunately, adding components to a system always comes with the natural drawbacks of impacting system weight limits and battery usage, which can then in turn impact the system performance. The other drawback for these components is simply that they require additional hardware that is, not standard with most UAV systems. Finally, the methods mentioned here can have a reduced communication range, as they can only communicate as far as their screens can be seen or their speakers heard clearly. Eventually, communications will likely incorporate some of these methods while also leveraging the motion of a UAV, which we will investigate throughout this work.

2.2.2 UAV Proxemics

Proxemics is "the study of how man unconsciously structures microspace—the distance between men in the conduct of daily transactions" as described by (Hall, 1963). It is another component that can be manipulated to assist or change the overall message attempting to be conveyed through a system (Baytas et al., 2019). Discusses the concept of understanding how distancing impacts interaction from a comprehensive view of social UAVs. Other works that have explored the impact of UAV distancing in interactions includes (Duncan and Murphy, 2013; Wojciechowska et al., 2019), who explored using vehicles at different heights (Acharya et al., 2017). Explored that effect using an untethered system in addition to comparing it a ground vehicle. The overall consensus across studies was that interactions within the social zone were preferred to the personal zone, which is in contrast to research with human-human or human-ground robot interactions.

2.2.3 Flight Paths

The benefit of using flight paths for communication from UAV to human has been briefly explored (Sharma et al., 2013). Explored using UAV paths to communicate affective information, and suggested direct vs. indirect use of space and changing the speed of the system are two components that have a direct effect on the valence. From their study they found that a direct quick motion gave higher valence (Szafir et al., 2014). Used flight to assist in communicating intended destination while the system also completed other goals. Overall, they found easing into the motion and arcing it made participants feel the motions were more natural and safe, which is also consistent with the idea that direct, quick motions increased participant valence in (Sharma et al., 2013).

2.3 Personality Model

To obtain a richer understanding of how people would respond to a UAV, it is also important to consider their projected emotion in relation to the UAV (Fong et al., 2003). Suggests that stereotype personalities can be created using immediate response emotions (Cauchard et al., 2016; Spadafora et al., 2016) explored this concept and presented an emotional model space for UAVs. Cauchard then also used these models to represent a full personality, or emotional state, such as Brave or Grumpy. These personalities, along with all individualized characteristics, could then be mapped based on varying speed, reaction time, altitude, and additional movement characteristics. Ultimately providing four different stereotypes of personality models that create the Emotional Model Space for UAVs. A few examples of these models include an Adventurer Hero or Anti-Social Drone. Understanding these categories allows us to better match a UAVs' action to expected action or scenario, in addition to some insight in how they may be perceived.

2.4 Affect, Attitude, and Perception

Interactions are biased by our previous experiences and interactions, but it can be difficult to know a participant's current affective state (and its impact on their study responses) without including a validated instrument. To understand the impact of a participant's previous experiences on their current interaction, questionnaires can provide this insight. One such instrument to better understand a participant's affective state and how it changes throughout the study is the Positive and Negative Affect Scale (PANAS) from (Watson et al., 1988). This questionnaire provides insight into how a participant is feeling that day compared to their normal state over the past week, and can be administered post-interaction to examine how the interaction impacts their state. Previous work by (Acharya et al., 2017) suggested that participants may have a higher negative affect after interacting with a UAV. The discomfort with the UAV was also supported by an increased distance in interaction when compared to a ground robot which did not result in an increased negative affect.

Another instrument, the Negative Attitude towards Robots Scale (NARS), has been suggested by (Riek et al., 2010) to impact a participant's ability to recognize humanoid motions, where participants with more negative attitudes were less able to recognize robot motions. NARS was introduced by (Nomura et al., 2004) and refined by (Syrdal et al., 2009). A participant's NARS score is calculated by averaging their values for three subcategories: Social/Future Implications, Emotional Attitudes, and Actual Interactions.

2.5 Crowdsourcing

Although running in-person studies may typically be preferred, online crowdsouring can be very useful in certain cases. There are a few cases where it may be more appropriate to use a crowdsourcing method. A few examples of these may include: when a large range of participants are needed, materials are targeted for refinement through many different proto-studies, or the work can be delegated into small tasks. Previous work by (Toris et al., 2014; Casler et al., 2013) have compared crowdsourced results to in-person and saw minimal to no difference in their results between the participants who came in person and those who completed tasks online.

3 EXPERIMENTAL METHODS AND DESIGN

This section describes experimental methods, materials, and design which are consistent across the phases of the studies to improve the readability of the article.

3.1 Pre and Post Interaction Surveys

Following a consent form, participants completed a demographic questionnaire, the first half of PANAS (based on their test condition, as listed in **Table 3**), and NARS. After the main task, they all completed a post-survey questionnaire consisting of questions about the study. If they completed PANAS prior to their task, they were asked to complete the second half of PANAS at this time.

3.2 Materials

For both Phase 0 studies an Ascending Technologies (AscTec) Hummingbird and Vicon motion capture system were used. For Phase 1–4 we used the DJI flamewheel F450, Pixhawk flight controller, and Vicon motion capture system.

3.3 MTurk

It is important to note the constraints on participants who were included in studies that were completed on Amazon Mechanical Turk (MTurk), which includes Phase 0 (Duncan et al., 2018), Phase 1, and Phase 3. Each participant's condition was dependent upon which of the mTurk task postings they selected. All tasks appeared the same to participants, so they had no insight into any differences and participants were excluded from future tasks once they participated in one. All participants were considered an MTurk "master," as determined by Amazon through analyzing worker performance over time. Also due to IRB restrictions from the GDPR privacy directive, none of the participants were allowed to be from the European Union.

Following any pre-interaction surveys, participants were redirected to a Google Form where they were asked to watch unique videos of a UAV flying in specific motions. The motions used for each phase are mentioned in their respective sections. Each video was 30 s in length, with repetitions added to reach the desired length if necessary. We used the Exhausted Drone template speed from (Cauchard et al., 2016) and the Anti-Social Drone altitude template to better compare to previous work. During a study participants would randomly be shown an attention check video that had a word displayed in the middle of it rather than simply showing a repeating motion. This check was placed to ensure participants were attending to the questions and watching the majority of the videos.

3.4 Motion Design

For the remaining participants, those in Phase 0 (Firestone et al., 2019) and Phase 4, they were presented with proposed states and asked to create motions to communicate those to others. In the case of Phase 0 (Firestone et al., 2019) this study was completed entirely in person. For Phase 4, the design and pre-interaction surveys were administered over Zoom and Google Forms, respectively. Following this they were asked to verbally describe and physically demonstrate their created motions using a small object (either a model drone in Phase 0 or an object roughly the size of a cell phone in Phase 4). The final component of the motion design study in either phase was to observe their drone flights in a Vicon motion capture space before completing the post-interaction survey.

4 PHASE 0

We now present the initial phase of the project, which includes two different studies. The first study explores label assignment at a high-level, looking for general agreement amongst participants. The second study explores user-defined flights created to convey the labels presented in the first study *via* an in-person setting.

4.1 Broad Agreement

Phase 0 (Duncan et al., 2018) involved 64 participants in total (43 Male, 21 Female). 56 identified as Americans, 2 as Chinese, 1 Korean, 1 Japanese, 1 Indian, 2 as "Other," and 1 did not respond. Each participant was paid 2 dollars and Amazon was paid 50 cents for recruitment. In the two alternative forced-choice (2AFC) task participants were given two labels, one of which was the expected label, and the other was a distractor chosen from a set of seven choices. In the seven alternative forced-choice (7AFC) task they were given all 7 of the options. Participants took 24.63 min (SD = 12.18) in the 2AFC task, and 26.15 min (SD = 12.29) in the 7AFC task.

The goal of this study was to understand if novice users showed broad agreement on the meaning of UAV gestures. To begin we looked to previously established protocols used for human gestures in (Krauss et al., 1991). Krauss looked to understand the level of participants' agreement by showing them a limited gesture set, followed by a request for them to apply a label from a limited set. Implementing this into a UAV gesture set began by exploring flight paths used by birds in nature and other biologically inspired behaviors, such as in (Arkin, 1998; Murphy, 2000).

4.1.1 Flight Path Labels

Labels were chosen based on flights that generally would require redirection, intervention, or awareness from either bystanders or operators. They were also chosen with the expectation that they would be well understood by novices due to their frequently observed use in other aircraft, being in general common system tasks, and similarity to other states in common technology (such as phones). The final consideration was choosing states that were domain independent, instead of focusing on applications (such as photography). Ultimately, the states chosen were: lost signal, lost sensor, draw attention, landing, missed goal, change position, and low battery.

4.1.2 Flight Path Selection

The original flight path selection was chosen to include motions that had steady periodic motion which could be created from sinusoid functions, to offer the ability to scale, and loop as needed. This in addition to drawing similarities to the biologically inspired avian flight paths originally identified by (Davis, 2000), lead to the eight cyclic motions of: Circle, Figure-8, Left-Right, Loop, Spiral, Swoop, Undulate, and Up-Down. Further details related to these choices and this work in general can be seen in (Duncan et al., 2018).

4.1.3 Results

The results in these studies were judged using a binomial test for 2AFC (compared to 50%) and a chi-squared test (compared to an even distribution) with $p < 0.01$; the resultant necessary agreement was 75% agreement in 2AFC and 34.4% agreement in 7AFC. In the 2AFC test the motions labeled with high agreement included Spiral (Landing, 87.5%), Figure 8 (Lost Sensor, 84.38%), and Swoop (Draw Attention, 75%). In the 7AFC test, 5 motions (3 unique from the first set) were significant at $p < 0.01$. Significant motions were: Circle (Draw Attention, 40.6%), Figure-8 (Change Position, 40.6%), Loop (Landing, 34.4%), Spiral (Landing, 59.4%), and Undulate (Draw Attention, 34.4%).

The full chi-squared values for the 7AFC are $\chi^2(6, N = 32) = 23$, $p < 0.001$ for Circle, $\chi^2(6, N = 32) = 22.6$, $p < 0.001$ for Figure 8, $\chi^2(6, N = 32) = 12.6$, $p = 0.049$ for Left-Right, $\chi^2(6, N = 32) = 19.4$, $p = 0.003$ for Loop, $\chi^2(6, N = 32) = 50.6$, $p < 0.001$ for Spiral, $\chi^2(6, N = 32) = 11.8$, $p = 0.066$ for Swoop, $\chi^2(6, N = 32) = 15.8$, $p = 0.01$ for Undulate, and $\chi^2(6, N = 32) = 9.4$, $p = 0.15$ for Up-Down.

Due to the number of chi-squared tests conducted, we are using the Bonferroni Correction to address possible effects found due to chance. Using this correction, our p-values will need to be

TABLE 1 | Taxonomy for UAV flight classification.

Taxonomy for user-designed flight paths		
Complexity	Simple	Single movement
	Compound	Collection of movements
Space	Direct	Focused approach to a point
	Indirect	Deviates from direct path
Cyclicity	Cyclic	Repeated motion (same path)
	Random	Singular flight path
Command	Roll	Left or right movement
	Pitch	Forward or back movement
	Yaw	Rotation
	Throttle	Up or down movement
Altitude	Increasing	Increase flight height
	Decreasing	Decrease flight height
	Variable	Increase and decrease
	Stable	No height change
Motion	Rectilinear	Only straight movement(s) and 90-degree turns
	Curvilinear	Only curved movement(s)
	Rotational	Only rotates
	Combinational	Combination of the above

below 0.0014 to still be considered significant at the same significance level, rather than below 0.01. Using this adjustment, only Circle, Figure 8, and Spiral are still considered significant.

Results overall showed a stronger understanding for Landing to be communicated by a spiraling path, and in general participants gravitated towards states that were less technical. This is shown from having stronger agreement for Draw Attention and Landing, and lower agreement for Lost Sensor. This finding shows support for the need of better refined labels that are commonly understood. Also due to the overall lack of strong agreement, this work suggested exploring open-ended responses and user generated flight paths. Finally, after initial observations, further research is needed to see if a negative NARS score suggests a decreased understanding of a UAV's motion, similar to the finding for humanoid robots in (Riek et al., 2010), or if a version of NARS should be revised to apply specifically to UAVs. The flight paths, the labels, and the application of NARS are investigated in the remainder of the paper.

4.2 Motion Elicitation

Phase 0 (Firestone et al., 2019) presented the same states from the earlier Phase 0 (Duncan et al., 2018) to twenty in-person participants (10 Male, 10 Female) who were local to the testing location in the United States. The cultural breakdown included 10 Americans, 2 Korean, 2 Indian, and 1 of each of Hispanic, Mexican, Austrian-American, Russian, European, and "other." As an incentive for participation they were each put into a drawing for a chance to win a $25 gift card. The seven states provided to participants were: Attract Attention, Sensor Lost, Low Battery, Signal Lost, Area of Interest, Missed Goal/Target, and Landing.

After eliciting a total of 140 gestures, a taxonomy was created to group the motions according to specific, common characteristics. This taxonomy encapsulates many different categorization/classification techniques. One of the most popular being the Laban Effort System best represented here by the complexity and space categories, from (Ruiz et al., 2011; Chi et al., 2000) respectively. Sharma et al. (2013) also previously used these two characteristics of the Laban Effort System to explore how they impacted people's perception of robotic motions, specifically flight paths. These categories are also well reflected within categories mentioned throughout (Venture and Kulić, 2019). This taxonomy is presented in **Table 1**.

The designed gestures were also grouped with common features according to the taxonomy, in addition to common motion characteristics. The most significant groupings were from Landing (thirteen people assign it as descending), Area of Interest and Missed Goal/Target (both had horizontal circles), and Low Battery (up-down motions).

A primary limitation of this work was the relative simplicity in a majority of the designed flight paths. This limitation was addressed in Phases 1–4 which followed to understand whether the difficulty in path creation was due to limited understanding of possible flight paths, difficulty with the initially defined states, or other limitations imposed by the experimental design.

5 EXPLORATION: PHASE 1

Based on the findings from the Phase 0 explorations into how participants would use a drone's motion to communicate, we embarked on an iterative approach in hopes of refining and collecting the different possible responses to drone motions. Further detail can be found in (Bevins and Duncan, 2021). A subset of motions demonstrating agreement from both studies in Phase 0 were presented to participants who were asked to respond to different questions about what they believed the drone was communicating and how they may respond.

5.1 Approach

The goals of this work were to validate the proposed videos for participant agreement, prototype questions for ability to elicit consistent responses, and understand the impact of asking multiple questions on participant responses. Throughout the study, other interesting considerations were encountered including the impact of pre- and post-questionnaires on the quality of participant responses, which is not central to the understanding of the motions, but is described further detail in (Bevins et al., 2020). The questions and processes developed were then used to better understand participants' expected perception and anticipated reaction to UAV flight paths. The full list of questions are presented in **Table 2**, with "Question Type" referring to the participant's anticipated response type. All of the questions were looking to obtain realistic answers to how participants' expect to perceive and/or react to a UAV's motion.

5.1.1 Question Variants

Three question types, each with two variations, were used in an attempt to obtain convergent responses with respect to participants' expected reactions. Gesture questions were expected to elicit participants' relation of the motion of the UAV to an action they have previously observed. Speech questions sought an anticipated

TABLE 2 | Study questions, with their anticipated response type, assigned number, and character length.

Question number	Question type	Question(s)	Characters
1	Speech	If you saw this drone in real life, what would it say to you?	61
2	Speech	If this drone could speak what would it tell you to do?	55
3	Gesture	What human gesture does this remind you of?	43
4	Gesture	If you had to replicate this movement with your head and/or body, what would you do?	84
5	Physical	If you were in the room with the robot, what would you do immediately following the robot's action?	99
6	Physical	If you were in the room with the robot, how would you respond immediately following the robot's action?	103

TABLE 3 | Question combinations for all test conditions within Phase 1.

Test condition	Question numbers asked	PANAS used
1 Speech	1	Yes
1 Speech	1	No
2 Speech	1, 2	Yes
1 Gesture	4	Yes
1 Gesture	4	No
2 Gesture	3, 4	Yes
1 Speech, 1 Gesture	1, 4	Yes
1 Speech, 1 Gesture	1, 4	No
1 Physical	5	Yes
1 Physical	6	Yes

verbal communication assigned to the UAV's motions. Physical questions sought to capture both speech and gesture aspects of the motion, in addition to a possible physical response. For this phase, participants would answer either 1 or 2 questions in a free-response method. The questions chosen are shown in **Table 2**. The full list of test conditions used, which question(s) were included in each condition, and whether that condition administered PANAS is shown in **Table 3**. Each line represents 8 participants.

5.2 Participants

Phase 1 had 80 participants in total (46 Male, 33 Female, 1 No Answer), with an age range of 24–68 (M = 38.6, SD = 10.7). Of the 80, 76 identified as American, 3 as Indian, and 1 as Chinese. The education levels were: high school (12), some college without a degree (17), college degree (46), and graduate-level degrees (4). Each participant was paid 4 dollars and Amazon was paid 1 dollar for recruitment. Across all of the conditions, participants took roughly 35 min.

5.2.1 PANAS

When examining the initial data that was collected from MTurk, the participants seemed to produce less diverse results towards the end of tasks (particularly those with single questions and double videos). To investigate the possible impact of participant fatigue, we removed the PANAS and additional videos during retests of selected conditions. All test conditions are listed within **Table 3**.

5.3 Videos

Participants were asked to watch 16 unique videos of a UAV flying in specific motions chosen and created from Phase 0 (Duncan et al., 2018) and Phase 0 (Firestone et al., 2019). This included all of the motions from Phase 0 (Duncan et al., 2018), complemented with a set of motions demonstrating the taxonomic differences and most popular flight paths from Phase 0 (Firestone et al., 2019). The base flight paths included: front-back, straight descend, descend and shift (descend then shift horizontally), diagonal descend, horizontal figure 8, horizontal circle, hover in place, left-right, plus sign, spiral, undulate, up-down, U-shape, vertical circle, X-shape, and yaw in place. Visualisations of these flight paths can be seen in **Figure 2**. Videos were each 30 s long, and if necessary repetitions of the flight were added to reach the desired length of the video. The paths were held constant for speed, around 0.5 m/s, and overall distance covered was held constant as much as possible. Depending on their condition, participants would see each video either once or twice. It was necessary to repeat a video set when they were asked two questions from the same category (two speech or two gesture). With each video they would receive either 1 or 2 questions. Each time they were asked to watch the entire video, but did have the capability of answering the question and proceeding, as there was not an attention check on every page.

5.4 Free Response Question Findings

An analysis of the question results sought to understand which question and/or question type produced the most actionable answers. Specifically, "actionable answers" referred to responses that indicated an intention for verbal or physical response to UAVs. The question type which proved most effective towards this goal was the "Physical" type. Since both questions of this type elicited similar results and only one was needed, we proceeded with "If you were in the room with the robot, how would you respond immediately following the robot's actions?" Further rationale for this decision is provided in (Bevins et al., 2020). For the purpose of the results presented here, and analysis within Phase 2, the responses were collapsed to be viewed as a single set. This choice was made due to the fact that responses in general were consistent enough for initial analysis, and seemingly more related to the flight path rather than the question.

6 REFINEMENT: PHASE 2

Through the methods described in this section, an analysis of the data from Phase 1 was conducted to determine which labels contained the most information, in addition to which question would be most effective. This section discusses that process and the steps taken for refinement in future phases.

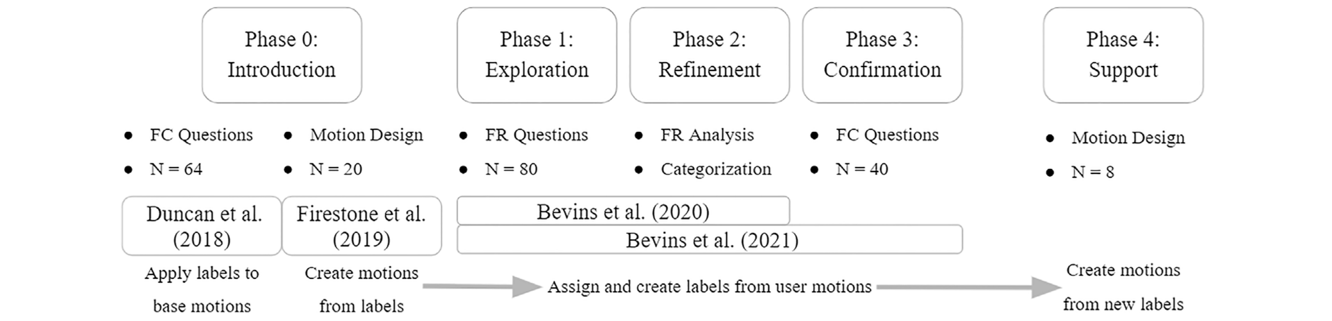


FIGURE 1 | Breakdown of Phases: initial introduction, an iterative creation/labeling exploration, and finally a supporting in-person study.

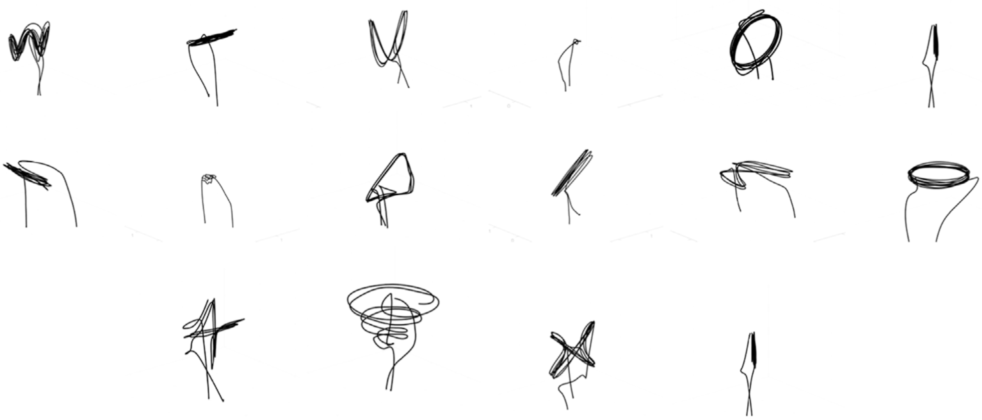


FIGURE 2 | Flight paths from top left to bottom right: undulate, left-right, U-shape, hover, vertical circle, up-down/descend, front-back, yaw, descend and shift, diagonal descend, horizontal figure 8, horizontal circle, plus, spiral, X-Shape, and up-down.

6.1 Frequency Analysis

For an initial understanding of the content of the responses, the 80 participants' responses from Phase 1 were roughly grouped based on the most commonly used words and general intent behind the words. An example of intent-based grouping would be how the words "stand" and "still" would both be sorted into a stare/observe type of category. From these methods we found 13 prominent categories that covered most of the expressed concepts which are listed in **Section 6.2**. Through this method we also found that many of the responses had participants describing the motion in some way, such as with "back" for front-back (25), "around" for yaw (20), and "side" for left-right (17). In addition to this, it was common to associate a motion with a human gesture that already exits, such as "nodding" for up-down (12) and "cross" for plus (6).

6.2 Category Formation

In addition to the states defined in the Frequency Analysis section, we incorporated categories that represented states such as delivery which are expected to be conveyed within UAV research. In most categories, multiple similar actions were combined to give raters a better sense of the types of responses that could be reasonably grouped together. The full list included:

- Follow/Follow a Path
- Blocked/Stop/Restricted/Do Not Pass
- Go Away/Back Away/Leave
- Move Towards/Approach
- Yes/Approval/Accept/Nodding
- No/Nagging
- Welcome/Hello
- Land/Falling/Lower
- Delivery
- Help
- Watch it/Caution/Slow Down/Investigate
- Stare/Hover/Look/Observe
- Power off

Two raters were asked to categorize the responses based on the provided categories. The raters were given instructions to choose a category only if they believed it appropriately fit, but to otherwise choose "Other." The raters ended up with kappa agreement scores over 0.93 for all of the categories, which shows excellent agreement (Landis and Koch, 1977). Overall, the goal for this method was to augment the findings from the frequency analysis to generate potential labels in future phases.

Looking into categories when responses are sorted by video provides a few further insights. 15/80 classified hover as "Stare/Hover/Look/Observe." 10/80 of front-back, 11/80 of horizontal circle, and 11/80 horizontal figure 8 were all classified as "Go Away/Back Away/Leave." 11/80 straight descend as "Land/Falling/Lower," 8/80 undulate responses sorted into "Blocked/Stop/Restricted/Do Not Pass," and finally 8/80 vertical circle as "Watch it/Caution/Slow Down/Investigate."

6.2.1 Forced Choice Definition

Following the raters' categorization, the categories were kept for inclusion if they showed high agreement and participant preference. Every category except for "Power Off" ended up being presented to the participants in Phase 3. From the categorization we also noticed that a second question for the participant would be beneficial to elicit answers in all of the categories, and provide more insight into how participants expected to respond. For this reason Question 1, "If you saw this drone in real life, what would it say to you?" was added to Question 6 when designing for Phase 3. The category options were then split across the two questions in an attempt to obtain convergent ideas between the two of them, while also allowing a comparison of the perceived communication with the intended reaction.

Five of the responses were appropriate choices for how participants plan to physically respond to a UAV: "Watch it/Look at it/Stare," "Investigate," "Follow it," "Move Away," and "Help it," in addition to an Other category.

The remaining categories were well suited for a speech category question because they helped communicate the states being conveyed to the person rather than showing a reaction to them. Since the responses being chosen here were states that could be communicated, a few of the categories were placed as response options in similar forms to both questions. All response options for Question 1 are: "To Follow It/Move Towards," "Do Not Follow/Do Not Pass/Restricted/Go Away," (DNF) "Yes/Approval," "No," "Welcome," "Landing," "Delivery," "Help," and "Caution" in addition to an added Other category.

7 CONFIRMATION: PHASE 3

Following the refinement phase, we were able to present a new set of participants with the newly generated labels and questions defined in Phase 2 from the data collected in Phase 1. Phase 3 consisted of 40 participants (19 Male, 20 Female, 1 No Answer), ranging in age from 25 to 57 (M = 39.1, SD = 8.1). Of the 40, 33 identified as American, 2 Chinese, 2 Indian, 1 Mexican, 1 Korean, and 1 did not answer. Each participant was presented with the 16 videos, for which they were asked to answer Questions 1 and 6 using the forced choice responses provided in **Section 6**.

A chi-squared test compared to an even distribution was used to find the statistically significant responses at $\alpha = 0.01$ with the participants from Phase 3, given a null hypothesis that all of the states should be chosen equally. All responses within **Table 4** (excluding yaw and the RFP rows) and in **Table 5** (excluding the RFP rows) reports significant results at the original given threshold. Similarly to the data presented in **Section 4.1.3**, due to the number of chi-squared tests conducted we need to address possible effects found due to chance. One way of addressing this is to use the Bonferroni Correction. Using this correction, our p-values will need to be below 0.000625 to still be considered significant, rather than below 0.01. Taking this into consideration, all responses within **Table 4** excluding Plus, Yaw, and the RFP rows are considered significant. All responses in **Table 5** excluding Undulate, Vertical Circle, and the RFP rows are considered significant. The effect sizes and p-values are provided in each section.

7.1 Perceived Communication

This section further discusses the results from the question "If you saw this drone in real life, what would it say to you?" In general, most participants assigned either DNF or "Landing."

The results presented in **Table 4**, suggest that participants would perceive a UAV to be blocking a path given large movements across the x-axis, with or without movement in the z-axis as well. Simpler motions with altitude changes were strongly associated with the intent to communicate "Landing." When the motions became more complex, incorporated a second direction (descend and shift), or additional axis of motion (spiral) it was not understood as clearly to mean "Landing" even though the dominant motion was within the z-axis.

There were a total of 640 responses, the breakdown of responses is represented by 25.7% responses for DNF, 15.9% for "Landing," about 13% for both "Caution" and "To Follow It/Move Towards," and 7.5% or less for each of the remaining categories. These values demonstrate that some categories are more likely to be chosen while others are either not well-defined or not anticipated to be associated with drone motions. These values are presented across all videos, but the distribution by video can be seen in **Tables 4**, **5**.

7.2 Anticipated Physical Response

Question 6, "If you were in the room with the robot, how would you respond immediately following the robot's actions?" the second question asked of participants saw a majority of responses for "Move Away" or "Watch it/Look at it/Stare," with the only significant deviation being front-back receiving an answer of "Follow it."

Some significant motion traits that appear when analyzing the responses for this question include "Watch it" responses having a key motion along the z-axis or not having movement along any of the axes. Vertical circle, descend and shift, yaw, up-down, plus, and diagonal descend all demonstrate this trend. Most of these motions also have a second highest choice of "Move Away," which likely explains the dissent within the straight descend and spiral paths. For these two specific motions, the popular choice was more evenly split between "Watch it" and "Move Away," of which the latter ultimately won out. The main takeaway from these results is that we can assume people would either watch or move away from vehicles that are relatively static or undergoing large altitude changes. "Follow it" was most prominent only with movements that were focused on the x-axis or x-y plane and approached closer to the participant, as shown with front-back

TABLE 4 | Q is Quantity of People providing that response, DoF is degrees of freedom, Sample Size is the total number of participants, and RFP refers to rotated flight paths with results only discussed in **Section 10**.

Motion	Say: Winning Response(s)	Q	DoF	Sample Size	Chi-Square Statistic	*p*-value	Cramer's V (effect size)
Undulate	Do not follow/Do not pass/restricted/go away (DNF)	14	9	40	52	$p < 0.0001$	0.360
Left-Right		14			50	$p < 0.0001$	
Horizontal Figure 8		14			54.5	$p < 0.0001$	
Horizontal Circle		15			42.5	$p < 0.0001$	
X-Shape		15			41	$p < 0.0001$	
U-Shape		13			33.5	$p = 0.0001$	
Hover		12			29.5	$p = 0.0005$	
Plus		11			23.5	$p = 0.0052$	
Vertical Circle		13			33	$p < 0.0001$	
Up-Down	Yes/Approval	15			39.5	$p < 0.0001$	
Spiral	Tie: DNF Tie: Landing	10			37	$p < 0.0001$	
Front-Back	To Follow It/Move Towards	23			124.5	$p < 0.0001$	
Yaw	Caution	7		32	13.5	$p = 0.1412$	0.528
Descend and Shift	Landing	21			92.5	$p < 0.0001$	
Diagonal Descend		23			112.5	$p < 0.0001$	
Straight Descend		22			103	$p < 0.0001$	
RFP: Undulate	DNF	5		8	22	$p = 0.0088$	0.429
RFP: Rotated Figure 8		4			16	$p = 0.0669$	
RFP: X-Shape	DNF/Landing	2			6	$p = 0.7399$	
RFP: U-Shape	DNF/Landing/Help	2			8	$p = 0.5341$	

TABLE 5 | Q is Quantity of People providing that response, DoF is degrees of freedom, Sample Size is the total number of participants, and RFP refers to rotated flight paths with results only discussed in **Section 10**.

Motion	Respond: Winning Response(s)	Q	DoF	Sample Size	Chi-Square Statistic	*p*-value	Cramer's V (effect size)
Undulate	Move Away	15	5	40	20.86	$p = 0.0008$	0.375
Left-Right		17			29.43	$p < 0.0001$	
Horizontal Figure 8		15			22.57	$p = 0.0004$	
Horizontal Circle		18			30.29	$p < 0.0001$	
X-Shape		18			32.00	$p < 0.0001$	
U-Shape		17			32.57	$p < 0.0001$	
Spiral		19			39.14	$p < 0.0001$	
Plus	Watch it/Look at it/Stare	15			29.71	$p < 0.0001$	
Vertical Circle		14			18.85	$p = 0.0020$	
Up-Down		16			25.43	$p = 0.0001$	
Hover	Tie: Watch it/Look at it/Stare Tie: Move Away	14			30.29	$p < 0.0001$	
Front-Back	Follow It	15			26.29	$p < 0.0001$	
Yaw	Watch it/Look at it/Stare	13		32	22.33	$p = 0.0004$	0.424
Descend and Shift		15			32.33	$p < 0.0001$	
Diagonal Descend		14			35.66	$p < 0.0001$	
Straight Descend	Move Away	12			25.00	$p = 0.0001$	
RFP: Undulate	Move Away	4		8	16.00	$p = 0.1562$	0.547
RFP: X-Shape		3			8.00	$p = 0.5494$	
RFP: Rotated Figure 8	Tie: Follow it Tie: Move Away	3			14.00	$p = 0.2206$	
RFP: U-Shape	Watch it/Look at it/Stare	3			10.00	$p = 0.4158$	

and horizontal figure 8. This led to an additional exploration of the RFP motions which is presented in **Section 10.2**.

Of the 640 responses, the breakdown of responses is represented by 36% of the responses were for "Move Away," 30% "Watch it," 18.8% "Investigate," 10.4% "Follow it," 4.5% "Help It" and Other was only chosen once for hover.

7.3 Free Response Within Forced Choice

With each of the questions participants had an "Other" option they could fill in if they felt none of the forced choice responses provided accurately portrayed their intentions. There were 13 total write-ins, accounting for a total of about 2% of the responses. None of the motions received more than 4 write-in answers. 12 in

TABLE 6 | Participants' chosen height of operation by state.

	Above head	Eye	Chest	Waist and below	Other
Do not follow/Go away	1	3	2	1	1
Watch it/Look at it	1	5	1	0	1
Investigate	2	1	4	1	0
Caution	2	1	4	1	0
Follow it/Move towards	2	1	5	0	0
Yes/Approval	3	3	2	0	0
Landing	1	2	1	3	1
Delivery	2	1	0	4	1

TABLE 7 | Participants' chosen speed of interaction by state.

	Fast	Average	Slow
Do Not Follow/Go Away	4	1	2
Watch it/Look at it	0	4	4
Investigate	1	3	4
Caution	1	3	4
Follow It/Move Towards	1	4	2
Yes/Approval	2	5	1
Landing	1	2	5
Delivery	0	4	4

total were written in for the perceived communication question from 8 different people, and only 1 answer was written in for the anticipated physical response question. Responses varied in content, but searching, confusion, and watching were popular among the write-ins.

8 PHASE 4

As an exploration to support the results from earlier phases, we presented 8 participants (6 Male, 2 Female) with the 8 communicative states used in Phase 3 to observe whether their motions would agree with the findings of Phase 3. Prior to participating these participants agreed to both an online and in-person session, so they are all local to the testing location in the United States.

Participants were asked to create flight paths to communicate states from Phase 3, similar to the methods of (Firestone et al., 2019), but over Zoom instead of in-person. Following this, they were expected to come in-person to view their flight paths on a real UAV, but for various reasons not all were able to complete the viewing portion of the study. This section also discusses the work of Phase 4 as compared to the other phases and related works.

8.1 Methods

After being greeted and consented, participants were asked to "please design an appropriate gesture, a flight path, for a drone to fly to communicate the state" for each of the states. After designing an appropriate gesture, they were asked to specify details about their motions, such as specific height, speed, and characteristics they would apply to their motions. They filled out a Google Form to answer all of these questions before verbally describing and physically demonstrating their motion using a small object of their choice (around the size of a cell phone).

8.1.1 Height

Participants were given the options of "Above Head," "Eye Level," "Chest Level," "Waist Level," "Knee Level," "Ground," and "Other" to associate with each motion. Due to low response choices, the options for waist, knee, and ground were grouped together for discussion. **Table 6** shows the full breakdown of heights chosen, sorted by state.

8.1.2 Speed

Participants were given the options "Fast," "Average," "Slow," and "Other" as options for their chosen speed. No further details about what concrete speed these choices entailed were provided. All eight participants answered this question for the majority of motions, but one chose "Other" for Do Not Follow, and another did not answer the question for Follow it. **Table 7** shows the full breakdown of speeds chosen, sorted by state.

8.1.3 Size and Space of UAV

Since participants created the gestures online they had no concept of where these motions would be used (i.e., indoor/outdoor) and thus how much space their UAV would have to fly. Some people

TABLE 8 | Motions created in Phase 4 classified according to the taxonomy and labeling from (Firestone et al., 2019).

State	Complexity	Space	Cyclicity	Command	Altitude	Motion
Do Not Follow/Go Away	Simple (5)	Direct (5)	Random (6)	Pitch (7)	Stable (5)	Rectilinear (7)
Watch it/Look at it	Simple (5)	Direct (4) Indirect (4)	Random (4) Cyclic (4)	Throttle (6) Roll (4)	Variable (3)	Rectilinear (4)
Investigate	Compound (5)	Indirect (5)	Random (5)	Roll (6) Pitch (5)	Stable (4)	Combinational (4)
Caution	Compound (6)	Indirect (6)	Random (5)	Roll (4)	Stable (5)	Rectilinear (6)
Follow it/move towards	Simple (5)	Direct (6)	Random (7)	Pitch (6)	Stable (5)	Rectilinear (7)
Yes/Approval	Compound (6)	Indirect (6)	Cyclic (5)	Throttle (7)	Variable (6)	Rectilinear (6)
Landing	Simple (4) Compound (4)	Direct (4) Indirect (4)	Random (7)	Throttle (7)	Decreasing (6)	Rectilinear (5)
Delivery	Simple (4) Compound (4)	Direct(5)	Random (8)	Roll (4) Pitch (4) Throttle (4)	Stable (4)	Rectilinear (3)

created motions that were either fully or slightly dependent upon the space that the UAV was flying in. For example, one person created motions that should go to the extremity of a person's view (fly as far as the operator could see it), or to the extremity of an available space (edges of a room). A more frequent response was to slightly scale up motions for a larger space/interaction area or larger UAV. The size of the UAV was also left open-ended, this appeared to cause some participants to think of the UAV as the size of the object they were holding.

8.1.4 Excluded Participants

In the early trials of running the Phase 4 study, and in response to the limitations identified from Phase 0 (Firestone et al., 2019), the experimenter showed brief demonstrations of possible flight characteristics. Due to anomolies in their responses this resulted in two participants, in addition to the eight described earlier, being excluded from the results, and analysis presented here in case they were unknowingly biased by the experimenter. During their task descriptions one participant was shown a circle and the other was shown line movements along axes. Both of these participants then showed these demonstrated characteristics consistently within their created flight paths. For the participant shown the circle 6/8 of their motions were categorized as curvilinear, and for the participant shown axis movement all of their motions were categorized as rectilinear. The remaining participants were not shown any example flight demonstrations. This exclusion raises significant concern on how seemingly small differences in experimental design with nascent technologies can unwittingly prime participant responses.

8.2 Results

We had participants recommend their preferred characteristics for an entire interaction space, including speed, height, and motion. The designed interactions section below provides a summary for each of the states, in addition to participants' speed and height characteristics.

8.2.1 Designed Flight Paths

Starting with "Do Not Follow/Go Away," five participants created different variations of a motion retreating from them, in addition to that two others chose small back-forth juts. This later motion is well reflected in the dominant speed trait, with fast being the most popular choice. For this motion it was also most common to place it around eye level.

For "Watch it/Look at it," two participants chose a yaw motion, for this motion we also see the first dynamic designs. With participants creating motions that either circled, created a diagonal line, or yaw towards the object of interest. There were also designs involving all three of those motion components that did not have a mentioned attachment to a specific area or object to observe. Five participants designed motions that they placed at eye level, and split their speed preference evenly between average and slow.

For "Investigate" the most dominant trait having movement along the x-y plane. Four of these motions involving a circle, three of which were horizontal. Most of them contained a line either moving left-right or front-back, but not both. Both Investigate and Caution were placed at a majority of chest level (with some eye level and above), and have a split for speed between average and slow. Looking deeper into the per-person breakdown shows that even though these two ended up with the same distribution, many of the participants chose different answers for each one (i.e., the same people didn't pick the same answers for both). The motions for "Caution" also don't have any curvilinear characteristics, and while three people designed a left-right motion, three more people also designed a vertical motion (up-down, vertical triangle) indicating further differences between the two states.

The "Follow it/Move Towards" motions, similar to the "Do Not Follow/Go Away," had six people create motions that moved away from the person. In these cases though the motions were more dynamic. A great example of this from one person is that they wanted the motion to make a line towards their destination with periodic yaws back towards the person. The remaining two suggested up-down changes. Overall the speed and height also show distinction between the two states. People here wanted the motion to be at chest level rather than eye, and chose an average speed rather than fast. This speed difference could indicate more of an offer for guidance (particularly paired with the yawing to ensure following) rather than fleeing in the earlier state.

For "Yes," all eight designed motions in the vertical plane, four of which were simply an up-down motion. Participants commonly noted that a reason for this was because it matches current human non-verbal communication in nodding or because it matches yes in sign language. These motions were placed at above head/eye levels with an average speed.

"Landing" also showed high agreement among participants, with six including a down motion, three of which were straight down. All of the motions involved the vertical plane, and two of them incorporated a yaw component. Every height category received placement, with slight majority going to waist and below, but there is much greater agreement that the motion should be slow in speed.

Finally, "Delivery" involved four participants designing an approach and three including a curved motion in various ways (curved approach, vertical circle, and "D" shape). Again placing the height at waist and below, and speeds of average or slow.

A couple of people mentioned when choosing motions placed below eye level that they wanted to be able to clearly see the UAV. One person described having it fly at this lower height gave them what felt like more control over the situation.

8.3 In-Person

Five of the eight participants were able to come in-person to view their created gestures performed by a DJI Flamewheel F450, the same vehicle that was used in the video recordings and in the same lab space used throughout the phases. After viewing their motions, they were asked if they would change anything. Most did not request any major changes to their originally designed motion, but all five mentioned changes they would make to at least one of their motions after viewing.

Typically these changes were in relation to the overall size of the motion, such making it larger or smaller. The amount that these motions were made larger was not consistent or a direct multiple of their small demonstration object to the size of the UAV. One participant designed motions that were originally proposed to be six inches in size, after viewing they determined it was not as clear as they desired. Another participant's motion was originally proposed to be approximately one foot and instead requested a change to six feet. One hypothesis for the large number of requested size changes was because participants may not have considered that a UAV, even in a highly controlled space with a Vicon system, has small perturbations while hovering. Because of this noise, the smaller motions were not usually large enough to create a clear distinction for their specific motion. Besides size changes, the only other change of note was when a participant requested to have the UAV move away from them rather than towards in all motions they designed with an approach (Watch it/Look at it, Investigate, Caution, and Delivery).

8.3.1 Added Modalities

At the end of their interaction each person was asked if they could add any modality to the UAV, what it would be. The responses were: Speaker/Sound x2, LED Panel x2 (green = good, red = bad/ stop) (green = follow me, yellow/orange = caution), and an on-board distance sensor to have the ability to act with a perception of the space around them.

8.4 Comparison to Previous Work

Phase 4 was run explicitly to compare to previous work in Phases 0–3 and to prior work by other colleagues working in this area. In this section, we will describe where this phase supports or contradicts work that has come before and then present areas that are well motivated for future studies.

8.4.1 Comparison to Phase 0 (Label Creation)

Only two of the states included in this phase were presented in Phase 0 (Duncan et al., 2018), "Landing" and "Draw Attention" which map to "Landing" and "Watch it/Look at it." The methods for Phase 4 are significantly different than in (Duncan et al., 2018), so any support is likely to be only *via* high-level flight path characteristics. Examining commonalities in responses between these works, we can see that all motions with a draw attention label (Circle, Loop, Swoop) are curvilinear, which we also see = in two of the eight motions designed for the "Watch it/Look at it" state. For "Landing," while one person did create a spiral in Phase 4 for landing, more common is a significant movement along the z-axis. From the similar characteristics found between the two works, we see very light support for Phase 0 (Duncan et al., 2018) results from Phase 4.

8.4.2 Comparison to Phase 0 (Gesture Elicitation)

The motions created by participants in Phase 4 were all categorized according to the taxonomy presented in Phase 0 (Firestone et al., 2019), and shown in **Table 8**.

Three states here are considered similar to those from Firestone (landing, investigate, and watch it/look at it). "Landing" is referred to by the same name here. In both of these studies throttle and decreasing altitude are considered significant, with weaker support for direct.

The second is area of interest, which we map to "Investigate" here. For both of these we see roll and pitch as significant commands. Four of the motions here are also curvilinear, supporting the motion finding.

Finally, the third is attract attention, which we map to "Watch it/Look at it." For this state, roll and throttle are the only characteristics that were considered significant for attract attention, and we see both of those represented here, with six out of eight motions containing throttle and four containing roll.

It should be noted the final two states do not perfectly map to states in (Firestone et al., 2019), but rather convey similar intents. In any case, the support seems reasonably strong for similarities in the structure of the designed motions indicating potential differences across states.

8.4.3 Comparison to Phase 3

Once again, when comparing across these phases, the expectation for Phase 4 to show support for Phase 3 findings would be based on high-level similarities between the created and selected motions. Notably, the same state options from Phase 3 are presented here, with only "Do Not Follow/Do Not Pass/ Restricted/Go Away" condensed down to "Do Not Follow/Go Away" differing.

This phase shows strong support for the idea that movement along the z-axis is distinguished as a characteristic of landing, with all participants having movement along the z-axis (six descend, two up-down). It also supports that "Follow it/Move Towards" should have large motions along the y-axis (six based on y-axis), although the motions are split (five and three) in terms of having an associated movement on the z-axis.

The recommendation to stay and watch a UAV was to minimize the amount of motion or have large altitude changes. Three of the states are presented as minimized motion, two yaw only and one circle defined as being only big enough to see movement. In addition to this, six participants include a throttle component, three of which were defined as moving a large amount (in these cases at least six feet). So we see some support for minimizing motion, but also overlap with the findings for landing. This is not unexpected and has been common across the studies where participants indicate an interest in watching a landing vehicle.

For "Yes" we again saw people associate up-down here, four provided basic up-down movements. In both phases, participants mentioned that this was because they associated the movement with nodding or yes in sign language.

In terms of the "Do Not Follow" state, we saw five of the participants design a motion that involved retreating (moving away) in some capacity, which is strange because many of the participants also designed a retreating motion to signify "Follow it/Move Towards." Phase 3 found it likely that movement along the x-axis would mean to not follow, so this phase does not support that finding. It should be noted that Phase 4 was the first to include differences in the speed of motions, so given this as an option it appears that participants may use speed to differentiate the meanings.

Another finding from Phase 3 indicated that complex motions should also result in participants moving away from an area.

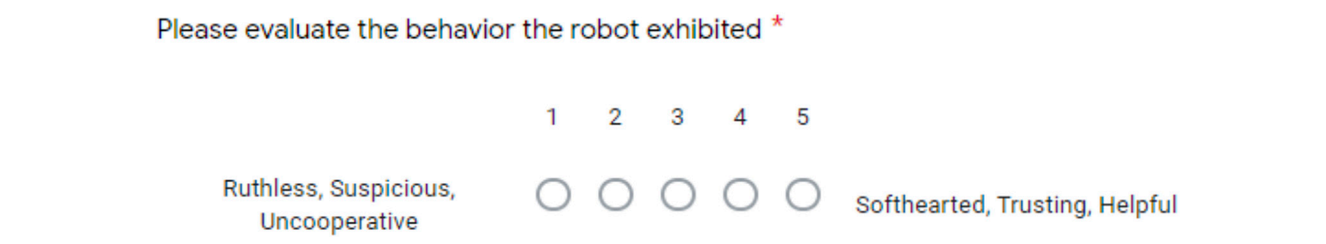


FIGURE 3 | Example of personality question as displayed to the participants in Phase 1 and 3.

These results are generally supported by the “Caution” state, which has six motions defined as compound.

Generally, we observe at least partial support for the findings in Phase 3 from the motions designed by participants in Phase 4. One notable exception is in “Do Not Follow/Move Away,” but this could also be due to the simplification of this state to exclude the idea of a restricted area after the conclusion of Phase 3.

9 CHARACTERISTICS

During the free response analysis in Phase 2, we quickly noticed people were responding with feelings within the responses regardless of whether we asked for it. Considering the findings from (Cauchard et al., 2016), we hoped to elicit similar personality traits, but were curious how participants would respond to flight paths when not varying the speed and height characteristics as in that work. Explicitly including this question also allowed us to investigate if different flight paths would elicit similar or different personalities.

We presented 2 independent raters, who were not participants, with the data from Phase 1 and asked them to attempt to categorize the responses into the emotional states from (Cauchard et al., 2016): Dopey/Sleepy/Sad, Grumpy/Shy, Happy/Brave, and Scared/Stealthy/Sneaky. They had high agreement (Kappa =0.63 and above) but indicated difficulty with the task. Feedback from raters indicated that they felt they were making a lot of assumptions by categorizing into these states, since it was typically inferred from an unrelated response. In future phases, we explicitly asked the questions to the participants and with the goal to gain complementary information regarding the states being selected. In regards to the raters, the overwhelmingly popular (by more than three times) category for both of them when sorting Phase 1 responses was Happy/Brave.

9.1 Personality Scale Definition

Modeled after (Spadafora et al., 2016; Cauchard et al., 2016) who presented the stereotypes of personality, each of the participants were given five scales they had to rank each of the videos using a 5 point Likert scale, pairing one extremity to the left side and the other to the right. The questions represented the “Big five” traits conveyed by two opposite poles: Openness to Experience, Conscientiousness, Agreeableness, Extraversion, and Neuroticism. **Figure 3** is a visual example of how this was presented to participants, and **Table 9** is the full list of presented characteristics.

9.2 Phase 3: Personality Characteristics

Most commonly participants classified the videos with Practical/Conforming, Organized/Disciplined, and Calm/Secure characteristics. According to (Cauchard et al., 2016) this meant that almost all of them would classify as brave, which Cauchard further goes on to classify as an *Adventurer Hero* Drone type, regardless of the motion depicted.

X-shape and undulate stand out as being more imaginative, disorganized, ruthless, and anxious in nature than the other motions. These four characteristics don’t perfectly match any of the models, but come closest to Sad, Dopey/Sleepy, and Scared, which closely resemble the *Exhausted* Drone. This is interesting because in (Cauchard et al., 2016) they involve significant altitude changes, and thus would be unlikely to be designed this way to convey such a state. The difference in perceived personality is also interesting given that both of these flight paths still elicited the most common forced choice responses of “Move Away” and DNF.

9.2.1 Personality Differences in Free Response vs. Forced Choice

Plus and Left-Right show opposite personalities when the participants were presented with free response options rather than forced choice. The responses for both motions showed significantly more imaginative traits assigned in free response, as categorized by the raters, and more practical in forced choice, as chosen by the participants. Again, this may be at least partially attributed to the experiment design as participants may be projecting their emotions onto what they see the UAV doing.

9.3 Phase 4: Personality Characteristics

During the online creation of participants’ motions, they were also asked to assign a UAV model to each state. Those responses are shown in **Table 10**.

Overall there is strong consensus for *Adventurer Hero*, which is the model most applicable to the results of Phase 3. Other states that diverge have converged to applicable archetypes, such as *Anti-Social* for “Do Not Follow/Go Away,” *Sneaky Spy* for “Investigate,” *Adventurer Hero* for “Follow it/Move Towards” and “Delivery,” and *Exhausted* for “Landing.” This lends support to both lines of work

TABLE 9 | Big five opposing characteristics presented as anchors to the Likert scale.

1	5
Practical, conforming, interested in routine	Imaginative, independent, interested in variety
Disorganized, careless, impulsive	Organized, careful, disciplined
Ruthless, suspicious, uncooperative	Softhearted, trusting, helpful
Retiring, sober, reserved	Sociable, fun-loving, affectionate
Anxious, insecure, self-pitying	Calm, secure, self-satisfied

TABLE 10 | Applied characteristics.

	Sneaky spy	Adventurer hero	Anti-social	Exhausted	Other
Do not follow/Go away	0	1	6	0	1
Watch it/Look at it	2	3	0	0	1
Investigate	4	2	1	1	0
Caution	0	2	2	2	2
Follow it/Move towards	2	4	1	1	0
Yes/Approval	1	3	0	1	3
Landing	1	1	1	5	0
Delivery	1	4	0	1	1

and calls for future studies explicitly linking the design characteristics from the designed motions here and the motion characteristics defined in (Cauchard et al., 2016).

There were still differences in the design of the motions when these motion characteristics were requested. For example, Cauchard places *Anti-Social* at about chest height and at an average speed. From our findings, "Follow it/Move Towards" motion had a large number of participants placing it at chest height with an average speed, but it is classified as *Adventurer Hero* by these participants. While none of the states have both a categorization of above head height and fast speed in this work, the closest resembling this is for "Yes/Approval," which participants also classify as *Adventurer Hero* and which matches the recommendations of Cauchard. The final set of parameters in Cauchard are for *Exhausted* personality profile. For this, the speed is slow and the altitude is best understood to be waist or below in this case. This best matches Delivery, which is also classified as *Adventurer Hero* by these participants.

9.4 Phase 4: In-Person Characteristics

The participants that came in-person to complete their study were presented with the same labels presented in **Table 9**, but on a scale of 1-6 instead. All eight states had a classification of practical, organized, softhearted, and calm when sorted as (1,2,3) and (4,5,6). Practical/conforming, organized/disciplined, and calm/secure were the same characteristics applied to the majority of videos in Phase 3. In addition to these four classifications, the only state that had a significant result on the Retiring/Sociable scale was "Retiring, Sober, Reserved" for "Landing," which had all five people classify it as a 3 (which is slightly agree on this scale). As before, this collection of characteristics doesn't map perfectly to any of the models, but of the options 3 of the 4 map to brave, happy, and shy. Happy and brave are condensed into the *Adventurer Hero* Drone, and shy falls under *Anti-Social*, regardless of the requested state. This could again be due to the lack of co-design in the personality characteristics and the motion of the UAV, so this is suggested for explicit inclusion in future work.

10 ADDITIONAL EXPLORATORY STUDIES

Throughout these studies opportunities were presented to gain additional knowledge about both state labels and the effect of the different axes of motion within the flight paths. Some of these opportunities were investigated *via* small proto-studies that were run in-between the larger studies to better inform their design. These additional investigations were not central to the narrative above, but do provide complimentary information for completeness.

10.1 State Elicitation

Between Phases 2 and 3, an additional sixteen participants (not included in any of the above studies) were asked for 3-5 states they believe a UAV should convey. Eight of these participants were also asked what information they believed a UAV should be able to communicate to those not involved in the UAV's operations. The question placement was counterbalanced between the beginning and the end of their study to see if participants provided more creative responses prior to applying given labels, or if they would provide the same states we provided if requested to provide states at the conclusion of the study. The placement of the request did not seem to have an effect overall. Regardless of placement, each of the participants submitted at least one of the states or labels that were included in the forced choice responses. The remaining portions of this study were not analyzed further due to poor responses. One lesson here, similar to that in the motion design and label creation categories is that creating prompts for participants which are open-

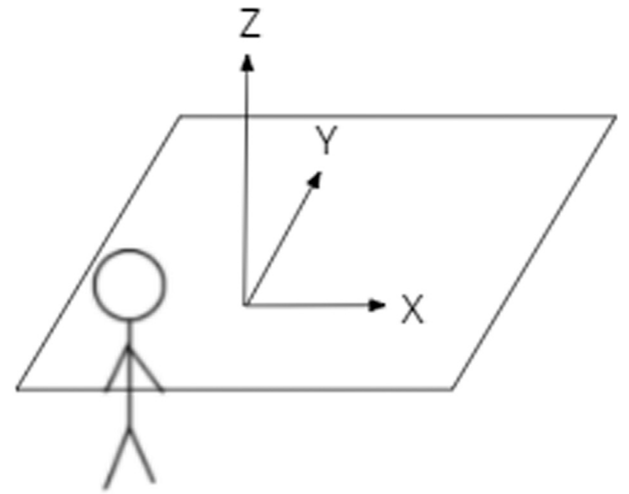


FIGURE 4 | Direction of axes of motion relative to participant.

ended enough to generate new ideas but narrow enough to provide overlap is a difficult endeavour.

10.2 Axis Investigation

After brief examination of the initial results of Phase 3 (the first 32/40 participants), we observed a seemingly consistent observation about the impact of the primary axis of motion. The initial observation was that motions moving mostly along the *x*-axis appeared as though they would elicit a blocked response, as demonstrated by all actions with the DNF choice were either significantly or solely on the *x*-axis. Whereas motions mostly on the *y*-axis seemed to rather encourage motion in that direction (to follow it), shown by front-back.

To test this observation, four of the motions that received the least amount of DNF categorizations from the first 32/40 participants in Phase 3 (front-back, straight descend, yaw, and diagonal descend) were replaced with four motions receiving the highest DNF classifications that had their primary axis of motion relocated from *x*-axis to *y*-axis (undulate, U-shape, X-shape, and horizontal figure 8). For these cases participants would then see both the original undulate on the x-z plane, in addition to an undulate on the y-z plane. This design was adopted to reduce any differences in this participant set. A visualization of the axes of motion relative to the participant is shown in **Figure 4**.

Ultimately there was not support for this initial observation within the exploratory dataset. The motions when rotated were still DNF, but we did observe a decrease in the intention to "Move Away" when compared to the earlier results. A different takeaway from these results is that it appears simplicity of the flight still holds a priority in effect, as with added complexity to the front-back motion we observed a change to a DNF state.

A noteworthy exception to the findings here is that horizontal figure 8, although initially classified DNF, when rotated received a tie for DNF and "Follow it" classifications. This could be due to the fact that this motion is unique from the others in that it moves a similar total x and y distance, with the distance on the *y*-axis from the participant being similar to that of the front-back motion. Another distinction this motion has from the other turned motions is a lack of motion on the *z*-axis. Overall this exploration is small and further study of these concepts would prove beneficial.

10.3 Phase 3 NARS Impact

The NARS questionnaire (Syrdal et al., 2009) contains questions asked on a Likert scale from 1 to 5, limiting participants' scores within a given category to an average between 1 and 5. A score below 2 is considered positive, and a score above 3 is considered negative. Values between 2 and 3 are considered neutral.

After reviewing the states that were being presented to participants in Phase 3, they fell within three natural groupings. The first grouping contains states that can be associated with a more positive connotation, while also being states that could be considered as welcoming movement towards the UAV. The second grouping was neutral states, or states that may invite the viewer to be stationary. Finally, the third grouping was negative sentiment states, otherwise viewed as states that encouraged the viewer to move away from the UAV or discouraged interactions.

- Positive/Move Towards: "To Follow It/Move Towards," "Yes," "Welcome," "Help," "Follow it," "Help it"
- Neutral/Stay: "Landing," "Delivery," "Watch it," "Investigate"
- Negative/Move Away: DNF, "No," "Caution," "Move Away"

In total, participants provided 32 responses to questions that were prompted with this set of responses (16 responses for each question, 2 questions per video). Observing the correlation between people's NARS scores and their chosen states, participants appeared more likely to choose a state from a given category based on whether they have a positive or negative NARS score. We observe that people with a NARS score classified as negative were more likely to pick negative states (mean:13.07, SD:4.7), and overall they were not as likely to choose one of the positive responses (mean:6.36, SD:3.4) $t(26) = 4.27$, $p = 0.0002$. Those with a positive NARS were likely to pick a positive state (mean:10, SD: 3.08) or negative state (mean:9.6, SD: 2.07) at about the same frequency $t(8) = 0.24$, $p = 0.815$. Both positive and negative NARS participants classified motions as one of the neutral options about 12 times on average [$t(17) = 0.12$, $p = 0.907$].

10.3.1 Personality Traits and NARS

Another correlation was between the NARS scores and the personality traits assigned to the motions. The 14 participants who had a negative NARS score were more likely to define the UAV as conveying practical, disorganized, ruthless, retiring, and anxious characteristics as seen in **Figure 5**. Whereas the 5 participants who had a positive NARS score generally selected the opposite traits (imaginative, organized, softhearted, sociable, and calm). The average of all 56 participants fell within the neutral values on all of the traits.

We test the null hypothesis that there is not a relation between NARS scores and chosen personality traits using t-tests. Using a

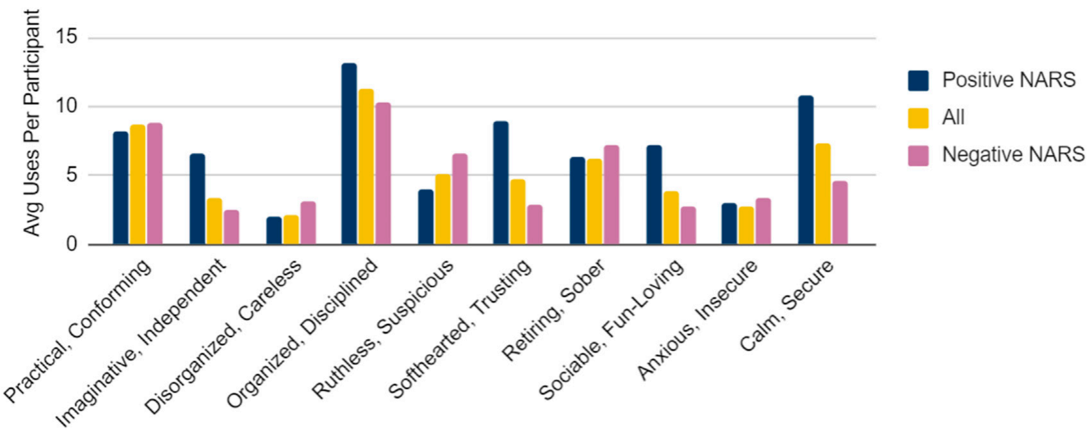


FIGURE 5 | Average number of times a personality category was chosen by a participant based on their NARS score. The upper bound for number of uses is 16 per participant.

t-test for 2 independent means we see that 4 of these results show significance, meaning that there is a correlation between participants' NARS score and categories chosen, with an *alpha* of 0.05. These four results are Imaginative $t(17) = 2.35$, $p = 0.031$, Softhearted $t(17) = 3.51$, $p = 0.003$, Sociable $t(17) = 2.24$, $p = 0.038$, and Calm $t(17) = 3.13$, $p = 0.006$.

There was no significant difference for practical $t(17) = -0.24$, $p = 0.815$, Disorganized $t(17) = -0.53$, $p = 0.601$, Organized $t(17) = 1.28$, $p = 0.216$, Ruthless $t(17) = -1.38$, $p = 0.187$, Retiring $t(17) = -0.33$, $p = 0.749$, and Anxious $t(17) = -0.18$, $p = 0.862$.

11 DISCUSSION

This work investigated how the general public would perceive and respond to communicative flight paths from UAVs through an iterative refinement of both flight paths and state labels. The limitations, implications, recommendations, and our reflections on this work will be presented in this section.

11.1 Limitations

A limitation for all phases of this study is that the flight controller used did not maintain precision control of the altitude of the UAV over time, because of this the paths were slightly varied based on the battery levels at the time of a specific flight. This is primarily a concern for the videos since these motions were intended to be held at exactly the same center position. This was also less of a concern in-person as the in-person flights were typically much shorter than the 30 s, and if a significant change was noticed in the flight controller's ability to hold the altitude the battery was just changed between demonstrations.

Although the biologically inspired motions chosen at the beginning of this research were expected to be culturally universal, the interpretations of the motions presented here are likely to be impacted by our participants' culture. This is related to how cultures interpret body movements differently, as discussed in (Sogon and Masutani, 1989; Kita, 2009). This idea is particularly supported by Andonova and Taylor (2012) which discusses the cultural associations with specifically head nodding/shaking. Although head-nodding means approval or "yes" in many countries, it does not mean this universally. For example, Bulgaria has a reversed response pattern, where a vertical head movement means "no" and horizontal head movement means "yes." Given the relatively limited representation of non-Americans in our studies this is important to note as it impacts the generalizability of these results to other cultures.

Another note is the differences in pay participants received throughout these phases. Pay for the tasks in Phase 0 was comparable to other similar tasks available on mTurk and to similar in-person studies at related universities at the time, but was ultimately determined to be too low. In addition to taking place in different years, this is why the pay was increased for future studies. It is possible that this may have impacted the quality of work provided by participants in Phase 0, as noted by (Litman et al., 2015).

Finally, the most significant limitation is that this work focuses exclusively on single-turn communication rather than multi-turn interactions such as in (Clinkenbeard, 2018; Csapo et al., 2012; Sidnell and Stivers, 2012). Future works should focus on more involved multi-turn interactions to better leverage the promise of this new communication modality.

11.1.1 Video

A limitation of the work is that Phases 0 (Duncan et al., 2018), Phase 1, and Phase 3 were all limited to remote viewing using video recordings. While an effective preliminary method, the main concern is that it likely impacted participants' ability to provide their true reaction, as there is almost always a difference between an expected reaction and a natural reaction. Another aspect of this work which might impacts participants' ability to accurately predict their true reaction could be the lack of previous UAV interactions, particularly in a social context. This would naturally increase the gap between their expected reaction and actual reaction, or interpretation. This concern is further reinforced by the fact that every person who came in-person during Phase 4 had at least one motion they wanted to modify.

Video use also eliminates the ability to explore varied UAV size and sound effect. While participants were asked to always have their sound enabled for the videos, there was no sound verification. This likely means that some participants had their sound off, or at a barely audible level. This is a problem that needs to be further explored in-person because of the high level of impact these factors can have on presence, fear, and interest in the machine.

The height, size, and speed of the recorded motions presented were held relatively constant in these studies, as opposed to being varied to elicit emotional responses as in (Cauchard et al., 2016). This is a limitation because varying these factors may allow exploration of additional communicative functions (rushing, thoughtful, contemplative, etc). This was not an oversight, but a priority for the study to reduce those factors and see what emotions or states were elicited specifically from the flight paths. The impact of these factors is briefly explored in Phase 4, but warrants further investigation.

In general, when presented with a forced choice option, most people agreed that these states appropriately conveyed the message they were looking for or at least did not care to write-in a response. While we cannot know for sure which of these is true, however since the results of Phase 4 generally confirm those of Phase 3, the categories seem appropriate choices.

11.1.2 Phase 4

A similar limitation (lack of choice and context) within the final phase was that the participants had to create the motions remotely over Zoom. As mentioned in further detail within **Section 8**, this reduced the fidelity of the participant interactions and raised questions from the participants about how and where this motion would be used. Some of this confusion could have been amplified because participants were purposefully not provided with any details about intended use or demonstrations so that they ideally create gestures that are able to be broadly applied. The danger in accidentally priming participants is discussed in more detail within that section, but also bears repeating here. Two excluded participants were each presented with curvilinear and rectilinear paths, respectively, as a demonstration before producing their own paths almost exclusively within those categories. Further investigation into appropriate context, demonstrations, and other priming mechanisms would be valuable when generating design characteristics for new technologies.

The size of the participant count within Phase 4 is also of note. There were limitations in having a larger participant pool participate in-person due to health and community regulations at the time of the study (in winter 2020). Thus the concept behind Phase 4 is support our larger online studies and provide possible paths for future work, rather than being a summative study to conclude the work.

11.2 Implications

We present an exploration into perceived communication, expected physical response, and emotional response to varied UAV flight paths. As a result of this, there are important practical implications discovered here for UAV developers and future researchers that may help to provide safe and knowledgeable interactions for the general public. This work indicates that people relatively easily associate motions applied in other situations onto UAVs, especially in the cases of Landing being conveyed with an altitude change, and a controlled up-down communicating "Yes." If a UAV begins to move away from someone at a lower height and slower speed, it is highly likely to be understood to follow it, especially if the motion is dynamic (periodic yaw to "look back" at the person, or clearly going in a specific direction). Because we also saw Do Not Follow have a retreating motion, the context added by the speed and height of interaction become highly important.

We were able to elicit different personalities, as described by (Cauchard et al., 2016), without varying the underlying flight characteristics and thus extending that work. One of the more significant deviations from (Cauchard et al., 2016) is that the undulate motion is used as a prototype of *Adventurer Hero*, but the participants here classify that motion as one of few to be *Exhausted*. Overall, participants classified almost all motions as Brave, and in turn the UAV as an *Adventurer Hero* type, which held across both Phase 3 and 4 and in spite of the UAV base characteristics being more closely aligned with those of the *Anti-Social* Drone and *Exhausted* Drone.

Overall, the work presented here builds and presents aspects in each new phase that support previous findings with at least a low level of confirmation by leveraging early findings as a starting point for exploration in this iterative process.

11.3 Recommendations

A major recommendation which has been presented in recent sections and in the discussion so far has been in the need for study on how to situate requests to participants in designing interactions with novel technologies without priming their responses and while still producing convergent ideas. The work presented here was a first step towards identifying common expected communications and underlying assumptions about the meaning of different flight paths, but still leaves many questions open regarding height, speed, and place of interaction. A challenge throughout this work has been establishing underlying mental models of UAV flight paths without priming those models towards specific path components (as discussed in **Section 8.1.4**).

Another recommendation is to explicitly bridge the work between (Cauchard et al., 2016; Bevins and Duncan, 2021) in order to apply the personality models to the designed flight paths and understand any changes in participant perception. Given how underexplored the area of human-UAV interaction has been, this work has converged in an interesting and exciting ways to build upon these lines of inquiry.

In a more fundamental sense, we have recommendations on flight paths, which include the complexity of motion, leveraging other motions within a culture, and the need to include the speed/height characteristics in future studies. From our results we found that complex motions frequently indicated an intention to move away from the UAV and/or area whereas simplifying or minimizing the motion would encourage them to stay and watch the UAV. Participants also associate motions applied in other situations well onto UAVs, especially in the cases of "Landing" being conveyed with an altitude change, and a controlled up-down communicating "Yes/Approval." If a UAV begins to move forward at a lower height and slower speed, it is highly likely to be understood to follow it, especially if the motion is dynamic (periodic yaw to "look back" at the person, or clearly going in a specific direction). Finally, as mentioned above, we note the need to have speed and height control to motivate a given context of interaction.

11.4 Reflection

An interesting result from the final Phase was the large amount of movement along the y-axis for the "Do Not Follow/Go Away" motions. At least one participant mentioned that if they were not supposed to follow the UAV then they would prefer that it depart the area (or at least their view). Contrast this to participants from Phase 3 where they perceived the "Go Away," as more similar to a guarding or protecting motion seen in a variety of communication scenarios (such as basketball guarding, a patrol team or dog). The dissonance between the two could be from a change in the state description where "Do Not Pass" and "Restricted" were removed as a simplification between Phase 3 and 4. While the authors assumed this change would have little to no effect on the responses, if this were a correct assumption, it could be assumed that a movement in-front of a person would give off a message that an area is blocked/to not approach, and to communicate not to follow is more associated with a speed and height than a particular motion (i.e., too fast and high). These types of findings can lead to a perceived brittleness in the studies conducted and the generalizability of the findings, however they could also be a testament to the difficulty in defining interactions with a new technology. Many of the findings were consistent across a 3-years, four phase study that was meant to both build and challenge its earlier results.

Throughout this work the most popular and recognizable characteristics of motion seem to frequently mirror an already recognized motion in a variety of domains. We see this represented most prominently with yes being associated with up-down and landing being associated with a straight descend, in addition to the note about guarding above. Everyday people, regardless of their design ability, have seemingly pulled these characteristics from interactions across human, object, or animal movement, and applied it as being effective in human-UAV communication. This is promising for future studies in this area, particularly in the open areas identified to provide support among disparate lines of research within this new field.

While this work has limitations, it extends the state-of-the-art in understanding how people interpret aerial vehicle motions to assist in informing how they may most effectively communicate messages (both via targeted motions and messages people are expecting to receive) to people using the most fundamental communication method in their flight paths. Future work is necessary to build upon the results shown here, but this work has taken a meaningful step towards bringing together previous work and understanding what people perceive about these systems.

11.5 Future Work

The most well-motivated future work described here is to merge the lines of research discussed in Phase 4 with those presented in (Cauchard et al., 2016) via a larger study either online or in-person. This would provide a richer set of flight paths with clear guidance on speed, height, and personality expectations to communicate specific states. This work could also be extended to understand how those flight path characteristics are impacted by the location of interactions and context inherent in the location changes (expectations for indoor versus outdoor, home versus public spaces, etc.).

As motivated in the limitations section, this work is focused exclusively on either one way communications or, at best, single-turn interactions. Future work would benefit from understanding how to leverage these into multi-party or at least multi-turn interactions. Recommended additional studies described below will also contribute to this improved understanding of a more involved or robust interaction.

Some specific limitations of the current work that could be addressed through additional studies include: adding context, adapting from designers in other areas, understanding the perception changes from in-person to online interactions, and the impact of additional communication modalities. It would be interesting to explore how flight paths vary when participants are given a specific scenario or use case to see how they adapt for each situation. As in other design work, it may prove beneficial to explore having animators, or dancers create the motions, as they are already trained in thinking about how to have people interpret motion that communicates messages. An extension of specifically Phase 3 would be to run the motions from Phase 3 in-person to see the full effects of being near the UAV as opposed to just viewing it online. Other factors to explore in the future that would compliment this work include adding light components, as mentioned by participants throughout, or changing the vehicle design.

Briefly addressed above and in Phase 4 would be further separation of categories combined here for simplicity, specifically splitting the"Do Not Follow/Do Not Pass/ Restricted/Go Away" category. While this category did provide general motivation, which was its purpose, it also appeared to be a catchall and may be better understood with separation of it into individual components. This was partially attempted by removing the restricted/go away between Phase 3 and Phase 4, it also appeared to lead to a large change in meaning and should have been split rather than simplified. Finally, a common note from in-person participants in Phase 4 was that they had imagined the motion would be more noticeable. This gap reinforces a rather simplistic understanding of UAV motion that we have (unsuccessfully) attempted to address in various phases. It is imperative to understand how to create a model of UAV flight, including the range of motion and inherent noise in the motion while not biasing the motions created by the participants. Perhaps this limitation could be addressed by providing a comprehensive explanation of typical UAV movement during describing the task or training participants on UAV flight characteristics, but there are concerns with priming responses when providing further details or demonstrations. This is a fundamental issue which needs to be addressed in future work to truly understand how to best leverage flight paths for communications.

12 CONCLUSION

Through this work we have been able to understand how participants would respond, both physically and emotionally, as well as better understand their perception of the messages naturally being conveyed within vehicle flight paths.

This work suggests that NARS can be an indicator of how a person may expect to respond and perceive the general sentiment of the message being conveyed. This work also indicates that people associate motions applied in other situations well onto UAVs. Especially in the cases of "Landing" being conveyed with an altitude change, and a controlled up-down communicating "Yes/Approval." If a UAV begins to move forward at a lower height and slower speed, it is highly likely to be understood to follow it, especially if the motion is dynamic (periodic yaw to "look back" at the person, or clearly going in a specific direction). Because we also saw "Do Not Follow/Go Away" have a retreating motion, it's highly important to note the need for speed and height situational control for proper context. Finally, flights crossing (moving along the x-axis) an area are likely to cause participants to avoid that area.

Finally, this work provides a roadmap to iteratively investigate the underlying communicative potential of new technologies while also raising significant questions about how to best elicit convergent states to communicate and common understanding of motion primatives. The discussion provided should be of keen interest to researchers investigating novel communication and to researchers in human-UAV interactions to understand where future work may have the most impact on bridging disparate investigations into this novel field.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Nebraska-Lincoln Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the National Science Foundation under IIS-1638099, IIS-1750750, CNS-1757908, IIS-1925368, and National Institute of Justice-GRF-STEM Fellowship 2016-R2-CX-0023.

ACKNOWLEDGMENTS

The authors thank Evan Beachly, Sebastian Elbaum, Carrick Detweiler, Urja Acharya, Siya Kunde, Justin Firestone, Rubi Quiñones, and Elizabeth Weber from the University of Nebraska-Lincoln for their assistance throughout the completion of this work.

REFERENCES

Acharya, U., Bevins, A., and Duncan, B. A. (2017). "Investigation of Human-Robot comfort with a Small Unmanned Aerial Vehicle Compared to a Ground Robot," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 2758–2765. doi:10.1109/iros.2017.8206104

Andonova, E., and Taylor, H. A. (2012). Nodding in Dis/agreement: a Tale of Two Cultures. *Cogn. Process.* 13, 79–82. doi:10.1007/s10339-012-0472-x

Arkin, R. (1998). *Behavior-based Robotics. Bradford Book*. MIT Press.

Baytas, M. A., Çay, D., Zhang, Y., Obaid, M., Yantaç, A. E., and Fjeld, M. (2019). "The Design of Social Drones: A Review of Studies on Autonomous Flyers in Inhabited Environments," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13.

Bevins, A., and Duncan, B. A. (2021). "Aerial Flight Paths for Communication: How Participants Perceive and Intend to Respond to Drone Movements," in Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 16–23.

Bevins, A., McPhaul, N., and Duncan, B. A. (2020). "Content Is king: Impact of Task Design for Eliciting Participant Agreement in Crowdsourcing for Hri," in International Conference on Social Robotics (Springer), 640–651. doi:10.1007/978-3-030-62056-1_53

Casler, K., Bickel, L., and Hackett, E. (2013). Separate but Equal? A Comparison of Participants and Data Gathered via Amazon's MTurk, Social media, and Face-To-Face Behavioral Testing. *Comput. Hum. Behav.* 29, 2156–2160. doi:10.1016/j.chb.2013.05.009

Cauchard, J. R., E, J. L., Zhai, K. Y., and Landay, J. A. (2015). "Drone & Me: an Exploration into Natural Human-Drone Interaction," in Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, 361–365.

Cauchard, J. R., Zhai, K. Y., Spadafora, M., and Landay, J. A. (2016). "Emotion Encoding in Human-Drone Interaction," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 263–270. doi:10.1109/hri.2016.7451761

Chi, D., Costa, M., Zhao, L., and Badler, N. (2000). "The Emote Model for Effort and Shape," in Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 173–182. doi:10.1145/344779.352172

Clinkenbeard, M. (2018). Multimodal Conversation Analysis and Usability Studies. *Commun. Des. Q. Rev.* 6, 103–113. doi:10.1145/3282665.3282675

Csapo, A., Gilmartin, E., Grizou, J., Han, J., Meena, R., Anastasiou, D., et al. (2012). "Multimodal Conversational Interaction with a Humanoid Robot," in 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom) (IEEE), 667–672. doi:10.1109/coginfocom.2012.6421935

Davis, J. W. (2000). *Categorical Organization and Machine Perception of Oscillatory Motion Patterns*. Ph.D. thesis (Cambridge, MA: Massachusetts Institute of Technology).

Duncan, B. A., Beachly, E., Bevins, A., Elbaum, S., and Detweiler, C. (2018). "Investigation of Communicative Flight Paths for Small Unmanned Aerial

Systems," in 2018 IEEE International Conference on Robotics and Automation (ICRA) (IEEE), 602–609.

Duncan, B. A., and Murphy, R. R. (2013). "Comfortable Approach Distance with Small Unmanned Aerial Vehicles," in 2013 IEEE RO-MAN (IEEE), 786–792. doi:10.1109/roman.2013.6628409

Firestone, J. W., Quiñones, R., and Duncan, B. A. (2019). "Learning from Users: an Elicitation Study and Taxonomy for Communicating Small Unmanned Aerial System States through Gestures," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 163–171. doi:10.1109/hri.2019.8673010

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. *Robotics autonomous Syst.* 42, 143–166. doi:10.1016/s0921-8890(02)00372-x

Funk, M. (2018). Human-drone Interaction. *Interactions* 25, 78–81. doi:10.1145/3194317

Hall, E. T. (1963). A System for the Notation of Proxemic Behavior. *Am. anthropologist* 65, 1003–1026. doi:10.1525/aa.1963.65.5.02a00020

Kita, S. (2009). Cross-cultural Variation of Speech-Accompanying Gesture: A Review. *Lang. Cogn. Process.* 24, 145–167. doi:10.1080/01690960802586188

Krauss, R. M., Morrel-Samuels, P., and Colasante, C. (1991). Do conversational Hand Gestures Communicate? *J. Personal. Soc. Psychol.* 61, 743–754. doi:10.1037/0022-3514.61.5.743

Landis, J. R., and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159–174. doi:10.2307/2529310

Litman, L., Robinson, J., and Rosenzweig, C. (2015). The Relationship between Motivation, Monetary Compensation, and Data Quality Among US- and India-based Workers on Mechanical Turk. *Behav. Res.* 47, 519–528. doi:10.3758/s13428-014-0483-x

Matrosov, M., Volkova, O., and Tsetserukou, D. (2016). "Lightair: a Novel System for Tangible Communication with Quadcopters Using Foot Gestures and Projected Image," in ACM SIGGRAPH 2016 Emerging Technologies, 1–2.

Murphy, R. R. (2000). *Introduction to Al Robotics*. Cambridge, USA: The Mit Press.

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2004). "Psychology in Human-Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots," in RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759 IEEE), 35–40.

Nozaki, H. (2014). "Flying Display: a Movable Display Pairing Projector and Screen in the Air," in CHI'14 Extended Abstracts on Human Factors in Computing Systems, 909–914.

Obaid, M., Kistler, F., Kasparavičiūtė, G., Yantaç, A. E., and Fjeld, M. (2016). "How Would You Gesture Navigate a Drone? a User-Centered Approach to Control a Drone," in Proceedings of the 20th International Academic Mindtrek Conference, 113–121.

Riek, L. D., Rabinowitch, T.-C., Bremner, P., Pipe, A. G., Fraser, M., and Robinson, P. (2010). "Cooperative Gestures: Effective Signaling for Humanoid Robots," in Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction (IEEE Press)), 61–68. doi:10.1109/hri.2010.5453266

Ruiz, J., Li, Y., and Lank, E. (2011). "User-defined Motion Gestures for mobile Interaction," in Proceedings of the SIGCHI conference on human factors in computing systems, 197–206. doi:10.1145/1978942.1978971

Scheible, J., Hoth, A., Saal, J., and Su, H. (2013). "Displaydrone: a Flying Robot Based Interactive Display," in Proceedings of the 2nd ACM International Symposium on Pervasive Displays, 49–54.

Sharma, M., Hildebrandt, D., Newman, G., Young, J. E., and Eskicioglu, R. (2013). "Communicating Affect via Flight Path Exploring Use of the Laban Effort System for Designing Affective Locomotion Paths," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 293–300. doi:10.1109/hri.2013.6483602

Sidnell, J., and Stivers, T. (2012). *The Handbook of Conversation Analysis*, 121. John Wiley & Sons.

Sogon, S., and Masutani, M. (1989). Identification of Emotion from Body Movements: A Cross-Cultural Study of Americans and Japanese. *Psychol. Rep.* 65, 35–46E. doi:10.2466/pr0.1989.65.1.35

Spadafora, M., Chahuneau, V., Martelaro, N., Sirkin, D., and Ju, W. (2016). "Designing the Behavior of Interactive Objects," in Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, 70–77. doi:10.1145/2839462.2839502

Syrdal, D. S., Dautenhahn, K., Koay, K. L., and Walters, M. L. (2009). "The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study," in Adaptive and emergent behaviour and complex systems.

Szafir, D., Mutlu, B., and Fong, T. (2015). "Communicating Directionality in Flying Robots," in 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 19–26. doi:10.1145/2696454.2696475

Szafir, D., Mutlu, B., and Fong, T. (2014). "Communication of Intent in Assistive Free Flyers," in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, 358–365. doi:10.1145/2559636.2559672

Toris, R., Kent, D., and Chernova, S. (2014). The Robot Management System: A Framework for Conducting Human-Robot Interaction Studies through Crowdsourcing. *J. Human-Robot Interaction* 3, 25. doi:10.5898/jhri/3.2.toris

Venture, G., and Kulić, D. (2019). Robot Expressive Motions. *J. Hum.-Robot Interact.* 8, 1–17. doi:10.1145/3344286

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: the Panas Scales. *J. Personal. Soc. Psychol.* 54, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wojciechowska, A., Frey, J., Sass, S., Shafir, R., and Cauchard, J. R. (2019). "Collocated Human-Drone Interaction: Methodology and Approach Strategy," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 172–181. doi:10.1109/hri.2019.8673127

Yamazaki, Y., Tamaki, M., Premachandra, C., Perera, C., Sumathipala, S., and Sudantha, B. (2019). "Victim Detection Using Uav with On-Board Voice Recognition System," in 2019 Third IEEE International Conference on Robotic Computing (IRC) (IEEE), 555–559. doi:10.1109/irc.2019.00114

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

REVIEW
published: 12 January 2022
doi: 10.3389/frobt.2021.721317

Evaluation of Socially-Aware Robot Navigation

Yuxiang Gao * *and Chien-Ming Huang*

Department of Computer Science, The Johns Hopkins University, Baltimore, MD, United States

As mobile robots are increasingly introduced into our daily lives, it grows ever more imperative that these robots navigate with and among people in a safe and socially acceptable manner, particularly in shared spaces. While research on enabling socially-aware robot navigation has expanded over the years, there are no agreed-upon evaluation protocols or benchmarks to allow for the systematic development and evaluation of socially-aware navigation. As an effort to aid more productive development and progress comparisons, in this paper we review the evaluation methods, scenarios, datasets, and metrics commonly used in previous socially-aware navigation research, discuss the limitations of existing evaluation protocols, and highlight research opportunities for advancing socially-aware robot navigation.

Keywords: socially-aware navigation, human-robot interaction, mobile robots, robot navigation, human-aware navigation

OPEN ACCESS

Edited by:
Adriana Tapus,
École Nationale Supérieure de Techniques Avancées, France

Reviewed by:
Marynel Vázquez,
Yale University, United States
Yomna Abdelrahman,
Munich University of the Federal Armed Forces, Germany

****Correspondence:***
Yuxiang Gao
yuxiang.gao@jhu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *06 June 2021*
Accepted: *06 December 2021*
Published: *12 January 2022*

Citation:
Gao Y and Huang C-M (2022) Evaluation of Socially-Aware Robot Navigation. Front. Robot. AI 8:721317. doi: 10.3389/frobt.2021.721317

1 INTRODUCTION

Fueled by advances in artificial intelligence (AI) technologies, mobile robots are realizing increased adoption in various delivery-based industries, from mail[1] and packages[2] to pizza.[3] Mobile robots designed for these consumer-facing services must not only navigate safely and efficiently to their destinations but also abide by social expectations as they move through human environments. For example, it is desirable for mobile robots to respect personal space (Althaus et al., 2004), avoid cutting through social groups (Katyal et al., 2021), move at a velocity that does not distress nearby pedestrians (Kato et al., 2015), and approach people from visible directions (Huang et al., 2014) while maintaining relevant social dynamics (Truong and Ngo, 2018). Research that investigates robot capabilities for navigating in human environments in an efficient, safe, and socially acceptable manner is commonly recognized as socially-aware navigation—also known as human-aware navigation (e.g., Kruse et al., 2013), socially compliant navigation (e.g., Kretzschmar et al., 2016), socially acceptable navigation (e.g., Shiomi et al., 2014), or socially competent navigation (e.g., Mavrogiannis et al., 2017).

While research on socially-aware navigation has expanded over the years (Kruse et al., 2013; Rios-Martinez et al., 2015; Charalampous et al., 2017; Pandey, 2017), there are no standard evaluation protocols—including methods, scenarios, datasets, and metrics—to benchmark research progress. Prior works on socially-aware robot navigation utilize a variety of evaluation protocols in custom settings, rendering comparisons of research results difficult. We argue that commonly agreed-upon evaluation protocols are key to fruitful progress, as observed in other research fields (e.g., computer

[1]Japan Post Co. piloted their mail delivery robot in Tokyo in October 2020.
[2]FedEx is currently developing the SameDay Bot for package delivery.
[3]Domino's launched delivery robots in Houston, TX, United States in April 2021.

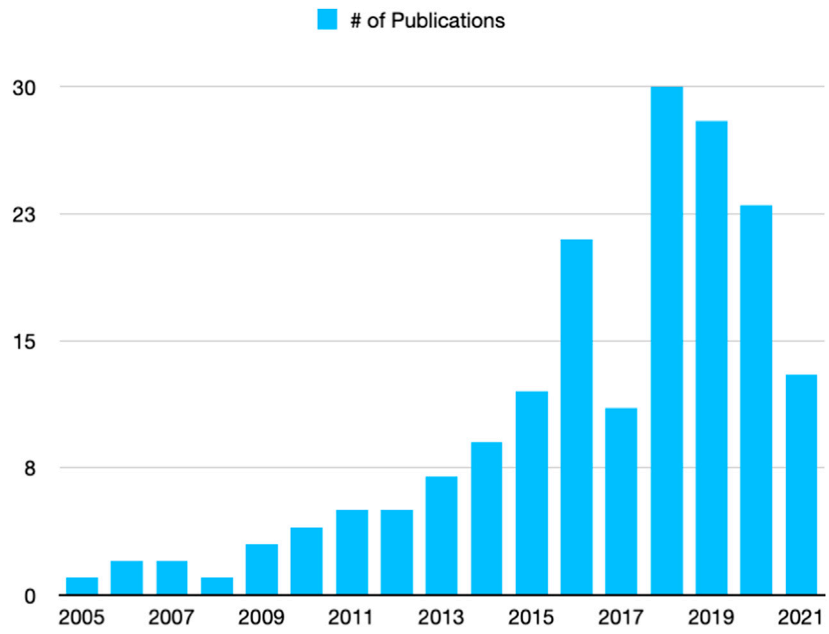


FIGURE 1 | Number of publications collected by year.

vision). As an effort to productively advance socially-aware navigation, in this paper we review commonly used evaluation methods, scenarios, datasets, and metrics in relevant prior research. We note that our review focuses on evaluation protocols rather than the algorithmic methods and systems that enable socially-aware navigation. We further note that socially-aware navigation is strongly related to an array of research topics, including human trajectory prediction, agent and crowd simulation, and robot navigation; some of the evaluation protocols reviewed in this paper may apply to these related research areas. Our review complements the recommendation for evaluation of embodied navigation suggested by Anderson et al. (2018) and can be consulted along with other general evaluation guidelines for human-robot interactions (Steinfeld et al., 2006; Young et al., 2011; Murphy and Schreckenghost, 2013).

The reminder of this paper is organized as follows. In **Section 3**, we present evaluation methods, scenarios, and datasets commonly used for evaluating socially-aware navigation. In **Section 4**, we review evaluation metrics and focus on the aspects of navigation performance, behavioral naturalness, human discomfort, and socialbility. We conclude this review with a discussion of limitations of existing evaluation protocols and opportunities for future research.

2 METHODOLOGY

Methodologically, this paper can be considered as a literature review—"*a literature review reviews published literature, implying that included materials possess some degree of permanence and, possibly, have been subject to a peer-review process. Generally, a literature review involves some process for identifying materials for potential inclusion—whether or not requiring a formal literature search—for selecting included materials, for synthesizing them in textual, tabular or graphical form and for making some analysis of their contribution or value*" (Grant and Booth, 2009). We focus on reviewing evaluation protocols for socially-aware robot navigation. While we did not follow the scoping process used for a systematic review, we identified materials (papers and datasets) for inclusion based on their relevance to the topic of socially-aware robot navigation and its evaluation methods. Specifically, we used keywords "socially-aware navigation,""socially-acceptable navigation," "human-aware navigation," or "crowd-aware navigation" when searching papers through ACM Digital Library, IEEE Xplore, and ScienceDirect. We additionally included some preprints from ArXiv through Google Scholar searches. This process yielded 188 papers in our initial search. Upon further reviewing the titles and abstracts of the papers, we removed 11 papers that did not address socially-aware robot navigation. The remaining 177 papers were published between 2005 and 2021 (**Figure 1**). A co-occurrence network of the keywords of the included papers is shown in **Figure 2**; the network illustrates three clusters that approximately represent topics related to human-robot interaction or social aspects of navigation (red), algorithmic methods for navigation (blue), and navigation systems (green). The co-occurrence network was automatically generated through Bibilometrix (Aria and Cuccurullo, 2017), a bibliometrics analysis tool, using Louvain algorithm. **Table 1** lists major venues where the 177 papers were published.

Upon collecting the 177 papers, we further reviewed the evaluation section of each paper and chose the studies that are representatives of the evaluation metrics, evaluation methods,

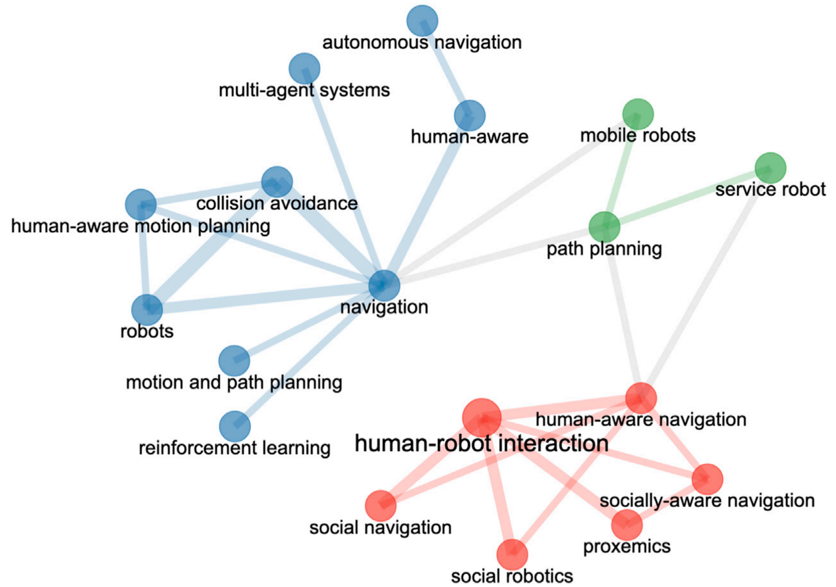


FIGURE 2 | Co-occurrence network of the keywords appeared in the collected publications. The keywords are clustered using Louvain algorithm. This graph is generated using Bibilometrix (Aria and Cuccurullo, 2017), a bibliometrics analysis tool.

TABLE 1 | Publication venues of the included 177 publications. Only venues that have more than five papers are listed.

Publication venues	#Of papers collected
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)	21
IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)	13
IEEE International Conference on Robotics and Automation (ICRA)	12
International Journal of Social Robotics	8
IEEE Robotics and Automation Letters (RA-L)	6
ACM/IEEE International Conference on Human-Robot Interaction (HRI)	5
Others	112

datasets, and test scenarios described in the next section. Through this process, we observed that many of the evaluation metrics were originated from related works on neighboring research topics such as human trajectory prediction, autonomous robot navigation, and crowd simulation. As a result, we include relevant works on these topics to better understand the development of the evaluation methods in our report and discussion below.

3 EVALUATION METHODS, SCENARIOS, AND DATASETS

In this section, we describe evaluation methods, scenarios, and datasets commonly used in socially-aware navigation research, some of which apply directly to the problems of human trajectory prediction, crowd simulation, and general robot navigation.

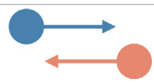
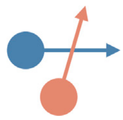
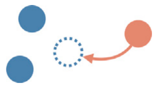
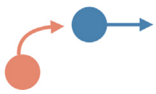
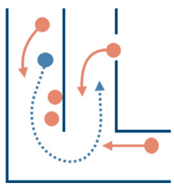
3.1 Evaluation Methods

Mavrogiannis et al. (2019) classified the evaluation methods into three categories: simulation study, experimental demonstration, and experimental study. In this review, we follow a similar but more granular classification based on the type, location, and goal of the evaluation methods. Specifically, we focus on four evaluation methods—case study, simulation and demonstration, laboratory study, and field study—regularly used in socially-aware navigation research. Each method has its own advantages and disadvantages and is often used at different stages of development.

3.1.1 Case Studies

Because navigating among people in human environments involves complex, rich interactions, it is common to break down socially-aware navigation into sets of primitive, routine navigational interactions such as passing and crossing (**Table 2**). As such, prior research has utilized case studies to illustrate robot capabilities in handling these common navigational interactions. Said case studies usually involve prescribed interaction behaviors (e.g., asking the test subjects to walk in a predetermined direction or behave as if they were walking together) and environmental configurations. For example, Pacchierotti et al. (2006) studied how a person and a robot may pass each other in a hallway environment; their study involved different human behaviors,

TABLE 2 | Scenarios commonly used in evaluating socially-aware navigation. The publications that employ each scenario in simulation or real-world settings are listed respectively.

Interaction type	Illustrations	Used in simulation	Used in real-world settings
Passing		Chen et al. (2017a); Vega et al. (2019b); Yang and Peters (2019); Randhavane et al. (2019); Pandey and Alami (2010); Pérez-D'Arpino et al. (2020)	Butler and Agah (2001); Pacchierotti et al. (2006); Kretzschmar et al. (2016); Okal and Arras (2016);
Crossing		Alahi et al. (2016); Chen et al. (2017a); Chen K. et al. (2019); Sui et al. (2019); Manso et al. (2019); Khambhaita and Alami (2020); Nishimura and Yonetani (2020); Daza et al. (2021)	Guzzi et al. (2013b); Kretzschmar et al. (2016); Johnson and Kuipers (2019); Mavrogiannis et al. (2019)
Overtaking		Kirby (2010); Pandey and Alami (2010); Anvari et al. (2015); Chen et al. (2017a)	Pandey and Alami (2010); Šochman and Hogg (2011); Robicquet et al. (2016); Yang and Peters (2019)
Approaching		Turner (1981); Sisbot et al. (2005); Truong and Ngo (2018); Johnson and Kuipers (2019); Truong and Ngo (2019)	Butler and Agah (2001); Satake et al. (2009); Kato et al. (2015); Truong and Ngo (2018); Joosse et al. (2021)
Following, leading, and accompanying		Ferrer et al. (2013b); Yao et al. (2019)	Ferrer et al. (2013a); Ferrer et al. (2013b); Ferrer et al. (2017); Du et al. (2019); Repiso et al. (2020)
Combined		Okal and Arras (2014); Okal and Arras (2016); Pandey (2017); Yang and Peters (2019)	Shiomi et al. (2014); Truong and Ngo (2018); Vega et al. (2019b)

such as moving at a constant speed or stopping in the middle of the hallway, and illustrated how the robot may respond to those behaviors. Similarly, Kretzschmar et al. (2016) reported a study demonstrating how their inverse reinforcement learning approach allowed a robotic wheelchair to pass two people walking together in a hallway without cutting through the group. Truong and Ngo (2017) presented an illustrative study comparing their proactive social motion model (PSMM) against the social force model (SFM) in four experimental settings and showed that their model yielded a more socially acceptable navigation scheme. Case studies can also be presented *via* simulation; Rios-martinez et al. (2013) used a set of predefined simulated configurations of human behaviors (e.g., moving around and interacting with each other) to illustrate their proposed method for reducing discomfort caused by robot movements.

3.1.2 Simulation and Demonstrations

Simulation experiments have been regularly utilized in recent years due to advances in reinforcement learning and data-driven approaches to socially-aware navigation (e.g., Chen C. et al., 2019; Li et al., 2019; Liu Y. et al., 2020). They are particularly useful for agile development and systematic benchmarking. Simulation experiments are typically supplemented by physical demonstrations to exhibit intended robot capabilities; the objective of these demonstrations is to illustrate that the proposed algorithmic methods work not only in simulated setups but also in the physical world with a real robot. For instance, Chen et al. (2020) first evaluated their method for crowd navigation in a simulated circle crossing scenario with five agents, after which they provided a demonstration of their method using a Pioneer robot interacting with human subjects. Katyal et al. (2020) and Liu L. et al. (2020) followed a similar method, including a simulation evaluation and a physical demonstration in their investigation of adaptive crowd navigation. Prior works that report this type of physical demonstration typically provide supplementary videos of the demonstrations (e.g., Jin et al., 2019).

Because of the popularity of simulation-based evaluation, an array of simulation platforms have been developed for robot navigation, ranging from simplistic 2D simulation [e.g., Stage (Gerkey et al., 2003) and CrowdNav (Chen C. et al., 2019), pedsimROS (Okal and Linder, 2013), MengeROS (Aroor et al., 2017)], to high-fidelity simulation leveraging existing physics and rendering engines [e.g., Webots,[4] Gibson (Xia et al., 2018), and AI2-THOR (Kolve et al., 2019)] and virtualized real environments [e.g., Matterport3D (Chang et al., 2017)]. Among these efforts, the following simulation platforms address socially-aware navigation specifically:

[4]https://www.cyberbotics.com

- *PedsimROS* (Okal and Linder, 2013) is a 2D simulator based on *Social Force Model* (SFM) (Helbing and Molnár, 1995). It is integrated with the ROS navigation stack and enables easy simulation of large crowds in real time.
- *MengeROS* (Aroor et al., 2017) is a 2D simulator for realistic crowd and robot simulation. It employs several backend algorithms for crowd simulation, such as *Optimal Reciprocal Collision Avoidance* (ORCA) (Van Den Berg et al., 2011), *Social Force Model* (SFM) (Helbing and Molnár, 1995), and PedVO (Curtis and Manocha, 2014).
- *CrowdNav* (Chen C. et al., 2019) is a 2D crowd and robot simulator that serves as a wrapper of OpenAI Gym (Brockman et al., 2016), which enables training and benchmarking of many reinforcement learning based algorithms.
- *SEAN-EP* (Tsoi et al., 2020) is an experimental platform for collecting human feedback on socially-aware navigation in online interactive simulations. In this web-based simulation environment, users can control a human avatar and interact with virtual robots. The platform allows for easy specification of navigation tasks and the distribution of questionnaires; it also supports simultaneous data collection from multiple participants and offloads the heavy computation of realistic simulation to cloud servers. Its web-based platform makes large-scale data collection from a diverse group of people possible.
- *SocNavBench* (Biswas et al., 2021) is another benchmark framework that aims to evaluate different socially-aware navigation methods with consistency and interpretability. As opposed to most simulation-based approaches where agent behaviors are generated from crowd simulation [e.g., using Optimal Reciprocal Collision Avoidance (ORCA) (Van Den Berg et al., 2011) or Social Force Model (SFM) (Helbing and Molnár, 1995)], human behaviors in SocNavBench are grounded in real-world datasets (i.e., UCY and ETH datasets) (**Section 3.3**). SocNavBench renders photorealistic scenes based on the trajectories recorded in these datasets and employs a set of evaluation metrics to measure path (e.g., path irregularity) and motion (e.g., average speed and energy) quality and safety (e.g., closest collision distance).
- The *CrowdBot* simulator (Grzeskowiak et al., 2021) is another benchmarking tool for socially-aware navigation that leverages the physics engine and rendering capabilities of Unity and the optimization-based Unified Microscopic Agent Navigation Simulator (UMANS) (van Toll et al., 2020) to drive the behaviors of pedestrians.

In addition to shared platforms for simulation-based evaluation, several online technical competitions have sought to benchmark socially-aware navigation. For instance, the *TrajNet++ Challenge*[5] focuses on trajectory prediction for crowded scenes and the *iGibson Challenge*[6] includes a social navigation task contextualized in indoor navigational interactions with human avatars.

3.1.3 Laboratory Studies

As opposed to case studies, which often involve prescribing human test subjects' behaviors (e.g., having them intentionally walk toward the test robot), laboratory studies utilize experimental tasks to stimulate people's natural behaviors and responses within specific contexts. Laboratory studies can be either controlled experiments or exploratory studies. Controlled experiments allow for statistical comparisons of navigation algorithms running on physical robots in semi-realistic environments; we note that controlled laboratory experiments contrast with simulation experiments, which lack the fidelity to represent real-world human-robot interactions. As an example, Mavrogiannis et al. (2019) designed an experimental task allowing three participants and a robot to move freely between six stations following a specified task procedure. A total of 105 participants were recruited for this experiment and a variety of objective and subjective metrics were collected to assess and compare three navigation strategies: Optimal Reciprocal Collision Avoidance (ORCA), Social Momentum (SM), and tele-operation. Additionally, Huang et al. (2014) evaluated how a humanoid robot may signal different levels of friendliness toward participants *via* movement behaviors—such as approach speed and direction of approach—in a mock museum setup.

Laboratory studies may also be exploratory, allowing researchers to gain early, prompt feedback from users without controlled experimentation. For instance, Bera et al. (2019) conducted an exploratory in-person lab study with 11 participants to investigate their perceptions of a robot's navigational behaviors in response to their assumed emotions.

3.1.4 Field Studies

While laboratory experiments allow for controlled comparisons, they bear reduced ecological validity; to address this limitation, field studies are used to explore people's interactions with robots in naturalistic environments. The pioneering tour guide robots RHINO (Burgard et al., 1998) and MINERVA (Thrun et al., 1999) were deployed in museums to study their collision avoidance behaviors and how people reacted to them. More recently, Satake et al. (2009) conducted a field deployment in which a mobile robot approached customers in a shopping mall to recommend shops; they explored different approach strategies and examined failed attempts. Similarly, Shiomi et al. (2014) investigated socially acceptable collision avoidance and tested their methods on a mobile robot deployed in a shopping mall for several hours with the objective of interacting with uninstructed pedestrians. Trautman et al. (2015) collected 488 runs of their experiment in a crowded cafeteria across 3 months to validate their algorithm. A benefit of deploying robots in the field is that they may reveal unexpected human behaviors; for instance, it was observed that young children "bully" a deployed mobile robot (e.g., intentionally blocking its way), which subsequently led to new research on how to recognize and avoid potential bullying behaviors in the field (Brščić et al., 2015). All in all, field studies

[5]https://www.aicrowd.com/challenges/trajnet-a-trajectory-forecasting-challenge
[6]http://svl.stanford.edu/igibson/challenge.html

TABLE 3 | Datasets used in socially-aware navigation.

Name	Year	# Of people	# Of scenes	Scene type	View type	Sensor type	Annotations	Publications
UCY (Lerner et al., 2007)	**2007**	**786**	**3**	**Outdoor**	**Top-down**	**Mono**	**Trajectories, Gaze**	Lerner et al. (2007); Alahi et al. (2016); Robicquet et al. (2016); Charalampous et al. (2016); Gupta et al. (2018); Vemula et al. (2018); Amirian et al. (2019); Yao et al. (2019); Kothari et al. (2020); Liu Y. et al. (2020); Biswas et al. (2021)
ETH (Pellegrini et al., 2009)	**2009**	**750**	**2**	**Outdoor**	**Top-down**	**Mono**	**Trajectories, Group Membership**	Pellegrini et al. (2009); Alahi et al. (2016); Charalampous et al. (2016); Robicquet et al. (2016); Gupta et al. (2018); Vemula et al. (2018); Amirian et al. (2019); Yao et al. (2019); Liu Y. et al. (2020); Kothari et al. (2020); Biswas et al. (2021)
Edinburgh Informatics Forum Pedestrian Database (EIPD) (Majecka, 2009)	**2009**	**95 ,998**	**1**	**Outdoor**	**Top-down**	**Mono**	**Trajectories**	Majecka (2009); Luber et al. (2012); Rudenko et al. (2017)
PETS2010	**2010**	**—**	**8**	**Outdoor**	**Surveillance**	**Mono**	**—**	Bandini et al. (2014); Bastani et al. (2015); Ristani and Tomasi (2015)
VIRAT (Oh et al., 2011)	**2011**	**4,021**	**11**	**Outdoor**	**Surveillance**	**Mono**	**Trajectories**	Oh et al. (2011); Vasquez (2016)
Town Centre (Benfold and Reid, 2011)	**2011**	**230**	**1**	**Outdoor**	**Surveillance**	**Mono**	**Bounding Boxes**	Benfold and Reid (2011); Ristani and Tomasi (2015); Le and Choi (2018)
Grand Central Station (Zhou et al., 2012)	**2012**	**12 ,600**	**1**	**Indoor**	**Surveillance**	**Mono**	**Trajectories**	Zhou et al. (2012); Gaydashenko et al. (2018)
CFF (Alahi et al., 2014)	**2014**	**42 million**	**1**	**Outdoor**	**Top-down**	**RGB-D**	**Trajectories, Bounding Boxes**	Alahi et al. (2014); Liu et al. (2020b); Kothari et al. (2020)
Stanford Drone Dataset (Robicquet et al., 2016)	**2016**	**11 ,216**	**8**	**Outdoor**	**Top-down**	**Mono**	**Trajectories**	Robicquet et al. (2016); Sadeghian et al. (2018a); Sadeghian et al. (2018b); Amirian et al. (2019); Li et al. (2020)
EgoMotion (Park et al., 2016)	**2016**	**—**	**26**	**Indoor**	**FPV**	**RGB-D**	**Bounding Boxes**	Park et al. (2016)
L-CAS (Yan et al., 2017)	**2017**	**6,140**	**1**	**Indoor**	**FPV**	**RGB-D**	**Trajectories**	Yan et al. (2017); Kothari et al. (2020); Liu Y. et al. (2020)
STRAND (Hawes et al., 2017)	**2017**	**—**	**1**	**Indoor**	**FPV**	**RGB-D**	**—**	Hawes et al. (2017)
WildTrack (Chavdarova et al., 2018)	**2018**	**9,518**	**7**	**Outdoor**	**Surveillance**	**Mono**	**Trajectories, Bounding Boxes**	Chavdarova et al. (2018); Liu Y. et al. (2020); Kothari et al. (2020)
JackRabbot Dataset (Martín-Martín et al., 2021)	**2019**	**260**	**—**	**Both**	**FPV**	**RGB-D**	**Trajectories, Bounding Boxes**	Martín-Martín et al. (2021)

are difficult to execute due to the unstructured, complex nature of real-world interactions—but are vital in evaluating socially-aware navigation and may offer insights that are otherwise impossible to discover in laboratory studies.

3.2 Primitive Scenarios

In this section, we describe common primitive scenarios found in the evaluation methods discussed in the previous section. **Table 2** summarizes primitive scenarios in evaluating socially-aware navigation by the nature of the interactions involved. These scenarios include:

- Passing: This scenario captures interactions in which two agents or groups are heading in opposite directions, usually in constrained spaces such as hallways or corridors, and need to change their respective courses to pass each other.
- Crossing: This scenario captures interactions in which two agents or groups cross paths in an open space; it also considers if one of the agents or groups is stationary. Common examples of this scenario are circle crossing, where all agents are initiated on points of a circle (e.g., Chen C. et al., 2019; Nishimura and Yonetani, 2020), and square crossing, where all agents are initiated on the corners of a square (e.g., Guzzi et al., 2013b).
- Overtaking: This scenario captures interactions in which two agents or groups are heading in the same direction and one of them overtakes or passes the other.
- Approaching: This scenario captures interactions in which a robot intends to approach or join a stationary or moving group or individual. This scenario is observed when a robot attempts to join a static conversational group (e.g., Truong and Ngo, 2018; Yang et al., 2020), initiate an interaction (e.g., Kato et al., 2015) or follow a moving social group (e.g., Yao et al., 2019).
- Following, leading, and accompanying: This scenario captures interactions in which a robot intends to join a moving group by following (e.g., Yao et al., 2019), leading (e.g., Chuang et al., 2018), or accompanying the group side-by-side (e.g., Ferrer et al., 2017; Repiso et al., 2020).

3.3 Datasets

Table 3 details a number of datasets of human movement that are regularly used in developing algorithms for and evaluating

TABLE 4 | Evaluation metrics for navigation performance.

Metric	Description
Path Efficiency	**The ratio between the distance between two waypoints and the length of the agent's actual path between those points**
Publications: Qian et al. (2010a); Guzzi et al. (2013b); Kruse et al. (2013); Stein et al. (2016); Sebastian et al. (2017); Honig et al. (2018); Mavrogiannis et al. (2018); Neggers et al. (2018); Johnson and Kuipers (2019); Mavrogiannis et al. (2019); Vasconez et al. (2019); Ahmadi et al. (2020); Batista et al. (2020); Chadalavada et al. (2020); Liang et al. (2020); Hacinecipoglu et al. (2020); Zhang et al. (2021)	
Success Rate	**Ratio of successful trials**
Publications: Burgard et al. (1998); Jin et al. (2019); Nishimura and Yonetani (2020); Guzzi et al. (2013b); Tsai and Oh (2020); Liang et al. (2020); Chen et al. (2019b,a); Honig et al. (2018); Kamezaki et al. (2020); Qian et al. (2010b); Samsani and Muhammad (2021); Sprute et al. (2019); Yao et al. (2021)	

socially-aware navigation systems. These datasets typically capture human movement in terms of trajectories or visual bounding boxes in various indoor and outdoor environments.

The datasets are used to train models for predicting pedestrian trajectories and for generating robot movement in the presence of pedestrians. In particular, they are commonly utilized in modern data-driven approaches to socially-aware navigation, such as deep learning methods (e.g., Alahi et al., 2016; Zhou et al., 2021; Kothari et al., 2020), reinforcement learning (e.g., Chen et al., 2017a; Li et al., 2019), and generative adversarial networks (GAN) (e.g., Gupta et al., 2018; Sadeghian et al., 2018a).

Datasets are also used to evaluate and benchmark the performance of socially-aware navigation (e.g., Biswas et al., 2021; Xia et al., 2018); for example, datasets ETH (Pellegrini et al., 2009) and UCY (Lerner et al., 2007) have been widely utilized in comparing navigation baselines (e.g., Sadeghian et al., 2018a; Bisagno et al., 2019; Gupta et al., 2018). One way to use the data of human trajectories in evaluation is to replace one of the human agents with the test robot agent and compare the robot's trajectory with the corresponding prerecorded human trajectory; various evaluation metrics described in the next section may be used to quantify the differences.

4 EVALUATION METRICS

In this section, we review common metrics used to evaluate socially-aware navigation. We begin by presenting metrics for assessing navigation performance in the presence of humans. We then review metrics for representing various aspects of social compliance; in particular, we focus on the three key aspects of social compliance in socially-aware navigation as proposed by Kruse et al. (2013): naturalness—capturing motion-level similarity between robots and people; discomfort—representing the level of annoyance, stress, or danger as induced by the presence of the robot; and sociability—encapsulating how well the robot follows the social norms expected by surrounding pedestrians.

4.1 Navigation Performance

In general, prior works used navigation efficiency (Guzzi et al., 2013a; Guzzi et al., 2013b; Mavrogiannis et al., 2018; Liang et al., 2020) and *success rate* (Burgard et al., 1998; Guzzi et al., 2013b; Jin et al., 2019; Liang et al., 2020; Nishimura and Yonetani, 2020; Tsai and Oh, 2020) to quantify the navigation performance of a robot. The common metrics for navigation performance are shown in **Table 4**.

4.1.1 Navigation Efficiency

We observed multiple measures of navigation efficiency in prior research, including path efficiency and relative throughput. Path efficiency is defined as the ratio of the distance of two waypoints to the length of the agent's actual path between those points (Mavrogiannis et al., 2019). Relative throughput (Guzzi et al., 2013b) is defined as the ratio of the number of targets the agent can reach if it ignores all collision and social constraints to the number of targets an agent can reach in an actual simulation. Both metrics calculate a ratio of performance under an ideal condition to performance under the actual condition, indicating the influences of interactions—either with people or the environment—on navigation efficiency. Other metrics for assessing efficiency include average velocity and mean time to goal (Liang et al., 2020).

4.1.2 Success Rate

In addition to the efficiency metrics discussed above, success rate is commonly used to quantify navigation performance in socially-aware navigation (Burgard et al., 1998; Guzzi et al., 2013b; Jin et al., 2019; Liang et al., 2020; Nishimura and Yonetani, 2020; Tsai and Oh, 2020). Success rate, or arrival rate, measures an agent's ability to reach its goal. When reporting success rate, it is also common to disclose the number of collisions and timeouts (e.g., Chen C. et al., 2019; Nishimura and Yonetani, 2020); a navigation trial is considered "timed out" if the agent cannot reach its goal within a specified time limit.

It is worth noting that success rate is highly dependent upon the environmental context and does not differentiate the quality of navigation between successful trials. As a result, weighted success rate metrics have been proposed to consider aspects of navigation efficiency, such as path length and completion time, while assessing success rate. These weighted metrics are single, summary metrics that represent navigation performance and can be particularly useful in reinforcement learning, which is a popular method used in recent works on robot navigation (Anderson et al., 2018; Yokoyama et al., 2021).

4.2 Behavioral Naturalness

Metrics related to naturalness focus on low-level behavioral patterns, i.e., how human-like and smooth robot movements

TABLE 5 | Evaluation metrics for naturalness.

Metric	Type		Description
	Similarity	Smoothness	
Average Displacement Error (ADE)	✓		**The average L_2 distance between the predicted trajectory and the human data**
Publications: Pellegrini et al. (2009); Alahi et al. (2016); Bera et al. (2017); Gupta et al. (2018); Anderson et al. (2019); Manso et al. (2019); Rudenko et al. (2019); Kothari et al. (2020); Zou et al. (2020); Hacinecipoglu et al. (2020); Zhou et al. (2021)			
Final Displacement Error (FDE)	✓		**The distance between the final destination in the prediction and the human data at the same time step**
Publications: Rudenko et al. (2019); Anderson et al. (2019); Kothari et al. (2020); Gupta et al. (2018); Manso et al. (2019); Bera et al. (2017); Pellegrini et al. (2009); Alahi et al. (2016); Zou et al. (2020); Zhou et al. (2021)			
Asymmetric Dynamic Time Warping	✓		**A trajectory measure that doesn't require both trajectories to have the same length**
Publications: Luber et al. (2012); Charalampous et al. (2016); Charalampous et al. (2017); Kostavelis et al. (2017); Avelino et al. (2021)			
Velocity and Acceleration		✓	**Basic dynamics measures**
Publications: Sisbot et al. (2005); Sisbot et al. (2007); Pandey and Alami (2010); Qian et al. (2010a); Qian et al. (2010b); Scandolo and Fraichard (2011); Kruse et al. (2012); Shiomi et al. (2014); Kollmitz et al. (2015); Kretzschmar et al. (2016); Truong et al. (2017); Truong and Ngo (2017); Claes and Tuyls (2018); Honig et al. (2018); Tail et al. (2018); Buchegger et al. (2019); Mavrogiannis et al. (2019); Papenmeier et al. (2019); Randhavane et al. (2019); Yoon et al. (2019); Zhong et al. (2019); Boldrer et al. (2020); Chadalavada et al. (2020); Fang et al. (2020); Hacinecipoglu et al. (2020); Ngo et al. (2020); Senft et al. (2020); Shiying et al. (2020); Gonon et al. (2021); Kivrak et al. (2021); Yao et al. (2021)			
Path Irregularity		✓	**The amount of unnecessary turning over the whole path**
Publications: Guzzi et al. (2013b); Mavrogiannis et al. (2018)			
Topological Complexity		✓	**Measures path entanglement to quantify encounters**
Publications: Kretzschmar et al. (2016); Mavrogiannis et al. (2018)			

are; measures of human similarity and path smoothness are also commonly used in human trajectory prediction research (Rudenko et al., 2019). A summary of the metrics for behavioral naturalness are shown in **Table 5**.

4.2.1 Movement Similarity

A common hypothesis in socially-aware navigation is that robots should possess navigational behaviors similar to humans' (Luber et al., 2012; Kruse et al., 2013). As a result, many prior works focus on developing and evaluating methods of producing robot trajectories that resemble those of humans under similar conditions. These prior works use a variety of measures—including displacement errors, dynamic time warping distance, and Hausdorff distance—to directly assess similarities between trajectories and end states in navigational performances.

Displacement Errors

Displacement errors are a family of metrics typically utilized in evaluating how well a predicted trajectory matches human trajectory data or a trajectory derived from other baseline methods. These metrics are widely used in pedestrian trajectory prediction research (Anderson et al., 2019; Rudenko et al., 2019; Kothari et al., 2020); they are also applied as evaluation metrics to assess the similarities between trajectories produced by navigation algorithms and by humans (Bera et al., 2017; Gupta et al., 2018; Manso et al., 2019; Kothari et al., 2020).

- Average Displacement Error (ADE) is the average L_2 distance between the predicted trajectory and the human data to which it is being compared. It was first used to evaluate trajectory similarity in socially-aware navigation by Pellegrini et al. (2009). As the nonlinear segments of a trajectory are where most of the social interactions between a robot and pedestrians occur (Alahi et al., 2016), ADE over these nonlinear portions provides a more specific metric for assessing human-robot navigational interaction.
- Final Displacement Error (FDE) is the distance between the final destination in the predicted trajectory and the human data at the same time step. It was proposed by Alahi et al. (2016) as a complement to ADE and nonlinear ADE.

Variations such as minimum, minimum over N, best-of-N, and top *n*% ADE and FDE are also employed by recent pedestrian trajectory prediction works (Anderson et al., 2019; Rudenko et al., 2019); these metrics distinguish the highest accuracy a prediction can achieve on human data, which is vital for trajectory prediction. However, accuracy is not a primary concern for socially-aware navigation research, which prioritizes learning general behavior patterns rather than generating exact matches of human trajectories; therefore, these variations are rarely applied to socially-aware navigation.

Dynamic Time Warping Distance

While displacement metrics are useful in characterizing overall trajectory similarities, they are inadequate in delineating the similarities between motion behaviors at different speeds; mismatched moving speeds are especially relevant to robot navigation as mobile robots have diverse form factors, resulting in widely varying velocities when compared to humans. To address this limitation, Luber et al. (2012) took a

TABLE 6 | Evaluation metrics for human discomfort.

Metric	Type			Description	Proposed in
	Spatial	Groups	Safety		
Personal space	✓		✓	**Spatial compliance for individuals**	Hall (1966)
Publications: Pacchierotti et al. (2006); Kessler et al. (2011); Scandolo and Fraichard (2011); Torta et al. (2013); Shiomi et al. (2014); Tomari et al. (2014); Talebpour et al. (2015); Kollmitz et al. (2015); Lindner (2016); Luo and Huang (2016); Kodagoda et al. (2016); Truong and Ngo (2016); Truong et al. (2016); Truong and Ngo (2018); Forer et al. (2018); Vega-Magro et al. (2018); Fei et al. (2019); Rajamohan et al. (2019); Randhavane et al. (2019); Bachiller et al. (2021); Banisetty et al. (2021); Batista et al. (2020); Fang et al. (2020); Fuse and Tokumaru (2020); Ngo et al. (2020); Shiying et al. (2020); Vega et al. (2020); Neggers et al. (2021)					
o/p/r-space	✓	✓		**Spatial compliance for static groups**	Kendon (2010)
Publications: Kessler et al. (2011); Scandolo and Fraichard (2011); Torta et al. (2013); Shiomi et al. (2014); Tomari et al. (2014); Kollmitz et al. (2015); Talebpour et al. (2015); Batista et al. (2020); Kodagoda et al. (2016); Lindner (2016); Luo and Huang (2016); Truong et al. (2016); Truong and Ngo (2016); Fei et al. (2019); Rajamohan et al. (2019); Randhavane et al. (2019); Fang et al. (2020); Forer et al. (2018); Truong and Ngo (2018); Vega-Magro et al. (2018); Fuse and Tokumaru (2020); Ngo et al. (2020); Shiying et al. (2020); Vega et al. (2020); Bachiller et al. (2021); Banisetty et al. (2021); Neggers et al. (2021)					
Social Force Model (SFM)	✓		✓	**Measures social compliance by artificial forces**	Helbing and Molnár (1995)
Publications: Šochman and Hogg (2011); Anvari et al. (2015); Huang et al. (2018); Kivrak and Kose (2018); Yang et al. (2019); Katyal et al. (2021)					
Extended social force model	✓	✓	✓	**Adds support for social groups to SFM**	Moussaïd et al. (2010)
Publications: Yang et al. (2019); Katyal et al. (2021)					

different approach by focusing on the fact that trajectories are time-series data bearing resemblance to spoken language; they proposed a modified version of Dynamic Time Warping (Sakoe and Chiba, 1978)—an algorithm commonly used for matching spoken-word sequences at varying speeds—to transform one trajectory into another *via* time re-scaling. A dynamic time warping distance can then be calculated to compare trajectories produced by agents moving at different velocities.

TABLE 7 | Interpersonal spaces as defined by Hall (1966).

Space name	Range	Function
Intimate space	$<.45m$	Intimate interactions
Personal space	0.45–1.2 m	Friendly interactions
Social space	1.2–3.6 m	Buffer zone for coexistence
Public space	$>3.6m$	Public interactions

4.2.2 Smoothness

The smoothness of both the geometric path and the motion profile of a robot are two important contributing factors to natural, safe navigation (Mavrogiannis et al., 2017; Mavrogiannis et al., 2018; Mavrogiannis et al., 2019). Not only are irregular paths and jittery movements inefficient, but they can also discomfort nearby pedestrians (Fraichard, 2007); therefore, it is critical to evaluate the smoothness of a robot's geometric path and motion profile in socially-aware navigation.

Path Irregularity

The smoothness of a trajectory can be characterized by the geometry of its path. For example, path irregularity (PI) (Guzzi et al., 2013b) measures the amount of unnecessary turning over the whole path a robot has traveled:

$$\mathrm{PI} = \sum_{\mathrm{Path}} \frac{\text{Robot Rotation} - \text{Min. Rotation Needed}}{\text{Unit Path Length}} \tag{1}$$

Topological Complexity

Prior research has also explored the use of the topological complexity index (Dynnikov and Wiest, 2007) to measure the level of entanglement in agents' paths (Mavrogiannis et al., 2018; Mavrogiannis et al., 2019). Greater path entanglement means that the agents are more likely to encounter each other during navigation, thereby inevitably forcing movement impact. Moreover, trajectories with simpler topological entanglements have been shown to be more legible (Mavrogiannis et al., 2018).

Motion Velocity and Acceleration

Velocity and acceleration are typically used to characterize motion profiles; a robot navigating in human environments is expected to keep a maximum velocity that allows it to reach the target while still maintaining a smooth acceleration profile. As an example, Mavrogiannis et al. (2019) used acceleration per segment and average energy per segment, where energy is the integral of squared velocity, to capture change in their robot's motion.

4.3 Human Discomfort

In this section, we present metrics used to measure human discomfort in socially-aware navigation. A summary of these metrics are shown in **Table 6**. We define discomfort as pedestrians' level of annoyance, stress, or danger caused by the robot's presence. Discomfort—either physical or psychological—is typically quantified by spatial models and subjective ratings (e.g., perceived safety).

4.3.1 Spatial Models

Spatial Models for Individuals

The impact of a mobile robot's navigational behavior on human comfort is difficult to quantify (Rios-martinez et al., 2013; Rios-Martinez et al., 2015; Kothari et al., 2020), as no universal "rules" are available for defining psychological comfort. Nevertheless, research suggests that the psychological comfort of humans is affected by interpersonal distance (Aiello, 1977; Baldassare, 1978; Greenberg et al., 1980). Proxemic theory (Hall, 1966) studies the function of the space an individual maintains for different social purposes in interpersonal interactions. According to Hall's observation, an individual's perceived personal space consists of several layers of concentric circles structured by their social functions, as presented in **Table 7**; however, according to Hall, most of his subjects were healthy business professionals from the northeastern seaboard of the United States. So these spaces may vary by culture and interaction context. Other representations—such as ovoids, concentric ellipses, and asymmetric shapes—have also been used to represent personal spaces and encode more complicated social rules (Rios-Martinez et al., 2015).

Among the four spaces laid out by Hall (1966), personal space is often used as the boundary of measuring perceived safety or social comfort—either as a no-go zone, where entering the space is counted as a violation of social comfort (Rios-martinez et al., 2013; Shiomi et al., 2014), or as the boundary of a potential function that assigns costs or penalties to robots entering that space (Amaoka et al., 2009; Truong and Ngo, 2018; Yang and Peters, 2019).

However, the circular representation of personal space as suggested by Hall (1966) is quite restrictive, as it does not adequately account for characteristics of human perception and motion. As a result, many works have explored different representations to consider face orientation (Amaoka et al., 2009; Truong and Ngo, 2016), approach pose (Truong and Ngo, 2018), and motion velocity (Helbing and Molnár, 1995; Truong and Ngo, 2016). Prior research has also leveraged empirical data from experiments to model complex and realistic uses of space (Gérin-Lajoie et al., 2008; Moussaïd et al., 2009). Most notably, the Social Force Model (SFM) (Helbing and Molnár, 1995), which has been widely used to simulate human navigation behavior in social contexts, represents the constraints of personal space as attractive or repulsive forces originating from each agent. Specifically, **Eq. 2** describes how an agent *i*'s behavior is driven by a combination of forces:

- $\overrightarrow{f_i}^{des}$: an attractive force that drives the agent to the desired goal.
- $\overrightarrow{f_i}^{obs}$: the repulsive forces from obstacles.
- $\sum_j \overrightarrow{f_{ij}}^{social}$: the sum of social repulsive forces from all other agents, *j*.

$$\frac{d\overrightarrow{v_i}}{dt} = \overrightarrow{f_i}^{des} + \overrightarrow{f_i}^{obs} + \sum_j \overrightarrow{f_{ij}}^{social} \tag{2}$$

Although SFM was designed for simulating crowd behavior, it has inspired metrics seeking to quantify social comfort in socially-aware navigation. For instance, repulsive forces from obstacles and nearby agents can be used to quantify violations of social comfort and indicate "panic" behaviors in emergencies (Mehran et al., 2009). Truong and Ngo (2018) proposed the Social Individual Index (SII) to measure the physical and psychological safety of an individual. Similarly, Robicquet et al. (2016) proposed the Social Sensitivity index, which uses potential functions to model how agents interact; high social sensitivity indicates that an agent will tend to avoid other agents.

Spatial Models for Groups

The aforementioned measures consider agents individually, but we must also consider that people interact socially in group settings. Social groups can be categorized into static and dynamic groups; static groups are groups of people standing closely together and engaging in conversations as commonly seen at social events, whereas dynamic groups are groups of people walking together toward shared destinations.

Static, conversational groups can be modeled using *f-formation* (Kendon, 2010). F-formation is the spatial arrangement that group members maintain in order to respect their communal interaction space, where *o-space* is the innermost space shared by group members and reserved for in-group interactions; *p-space* surrounds the *o-space* and is the space in which members stand; and *r-space* is the outermost space separating the group from the outer world. Similar to individual discomfort, discomfort caused by a robot to a group may be measured by the robot's invasion into either the *r-space* or the *o-space*, based on the f-formation of the group (Mead et al., 2011; Rios-martinez et al., 2013; Ferrer et al., 2017).

It is commonly observed that people walk together in dynamic social groups (Federici et al., 2012; Ge et al., 2012). In addition, individual people tend to stay away from social groups when walking (Efran and Cheyne, 1973; Knowles et al., 1976; Moussaïd et al., 2010). A mobile robot deployed in human environments must know how to behave around human groups by observing such inherent etiquette. To simulate dynamic social groups, Moussaïd et al. (2010) proposed the Extended Social Force Model (ESFM).[7] As shown in **Eq. 3**, ESFM adds a new group term $\overrightarrow{f_i}^{group}$ that dictates intra-group dynamics to the original SFM. The group term, as defined by **Eq. 4**, is the summation of three forces: a cohesive force that defines attractions between group members $\overrightarrow{f_i}^{att}$; a repulsive force between group members $\overrightarrow{f_i}^{rep}$; and a gaze force $\overrightarrow{f_i}^{gaze}$ that aligns each agent with the center-of-mass of the social group, factoring in head orientation to simulate in-group social interactions.

$$\frac{d\overrightarrow{v_i}}{dt} = \overrightarrow{f_i}^{des} + \overrightarrow{f_i}^{obs} + \sum_j \overrightarrow{f_{ij}}^{social} + \overrightarrow{f_i}^{group} \tag{3}$$

$$\overrightarrow{f_i}^{group} = \overrightarrow{f_i}^{att} + \overrightarrow{f_i}^{reb} + \overrightarrow{f_i}^{gaze} \tag{4}$$

Similar to spatial models for individuals, spatial models for groups can be used to approximate discomfort in group interactions. As an example, to evaluate a robot's social compliance as a group member when accompanying humans, Ferrer et al. (2017) proposed a quantitative metric based on the robot's position in relation to the human members, accounting for whether or not the robot was in the field of view of the human members and the distances between group members.

4.3.2 Physical Safety

Safety is the preeminent concern in socially-aware navigation. At the most basic level, navigational safety amounts to collision avoidance: a

[7]Our implementation of ESFM—https://github.com/yuxiang-gao/PySocialForce

mobile robot should not have any physical contact—intentional or otherwise—with any human being. Metrics based on collision count or violation count are commonly used in simulated environments and in some robot-only experiments. For example, Liu L. et al. (2020) used the number of collisions with agents within and without the test agent's field of view, along with success rate, as the main evaluation metrics in conducting their assessment of their deep reinforcement learning based navigation algorithm in simulation. Guzzi et al. (2013b) used small-scale robots in physical experiments, allowing them to use collision count as one of their main metrics in evaluating the impact of safety margin size.

While they are arguably the most straightforward methods of measuring navigational safety violations, collision and violation counts are neither practical nor ethical to use in real-world experiments and deployments involving humans, as collisions present potential harm to the participants. Consequently, safety violations should be approximated by invasions of defined safety zones. A safety zone is typically derived from the proxemics theory proposed by Hall (1966), wherein the personal space—ranging from 0.45 to 1.2 m in Western culture—is used to measure how well a mobile robot maintains the physical safety of nearby human pedestrians (e.g., Vega et al., 2019b). Variations on safety zones are frequently used in prior works; for example, the Collision Index (CI) (Truong and Ngo, 2016), or Social Individual Index (SII) (Truong and Ngo, 2018), is a distance-based metric for capturing the violation of personal space. The index is specified in **Eq. 5**, where (x_i^p, y_i^p) is the position of the ith pedestrian p_i, (x^r, y^r) is the position of the robot, and σ_0^{px} and σ_0^{py} are the standard deviations of the personal space, empirically set to the value of 0.28:

$$\mathrm{CI} = \max_{i=1:\,N} \exp\left(-\left(\frac{x^r - x_i^p}{\sqrt{2}\,\sigma_0^{px}} + \frac{y^r - y_i^p}{\sqrt{2}\,\sigma_0^{py}}\right)\right) \tag{5}$$

In the original definition of the index (Truong and Ngo, 2016), the standard deviations are the same for both directions ($\sigma_0^{px} = \sigma_0^{py}$), thus assuming that personal space is a perfect circle. However, as we discussed earlier, additional representations of personal space have been proposed to capture nuanced social rules, cultural influences, and specific situations (Rios-Martinez et al., 2015); therefore, this index may be adapted to account for different cultures, types of relationships, and interaction contexts by modifying the standard deviations. As another example of custom safety zones, Jin et al. (2019) defined the ego-safety zone as a circular space around an agent, analogous to the personal space, and the social-safety zone as a rectangular region stretching along an agent's current moving direction.

4.3.3 Psychological Safety

In addition to preserving physical safety, it is important to evaluate the effects of socially-aware navigation on psychological safety. Preserving psychological safety, or sometimes referred to as perceived safety, involves ensuring a stress-free and comfortable interaction (Lasota et al., 2017). Although they may not physically endanger a person, a mobile robot's navigational behaviors (e.g., how they approach and pass a person) may yet induce feelings of discomfort or stress (Butler and Agah, 2001). Consider a situation in which a mobile robot moves rapidly toward a person and only changes its moving direction right before the imminent collision; while the robot does not make direct physical contact with the person, its navigational behavior is still likely to cause them significant stress.

A common method of assessing people's perceived psychological safety is through questionnaires. Butler and Agah (2001) asked participants to rate their comfort from 1 to 5 (with 1 being very uncomfortable and 5 being very comfortable) under different experimental conditions, including varying robot speed, distance from the human subject, and approach patterns. Similarly, Shiomi et al. (2014) used a survey to assess people's experiences interacting with a deployed mobile robot during a field study; specifically, the inquiry focused on three aspects: whether the interaction was free from obstruction, whether the person could maintain their preferred velocity in the presence of the robot, and their overall impression of the encounter.

Several established questionnaires designed for social robotics research already include questions regarding psychological safety. For example, the Godspeed questionnaire (Bartneck et al., 2008) has a sub-scale, perceived safety, comprised of questions related to subjects' relaxed/anxious, calm/agitated, and surprised/quiescent emotional states. The Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017), based on the Godspeed questionnaire, measures people's perception and judgement of the robots' social attributes, including warmth, competence, and discomfort. The BEHAVE-II instrument (Joosse et al., 2013) includes a set of behavioral metrics that measure human responses to a robot's behavior; some of the metrics were specifically designed to gauge the discomfort caused by a robot's approach behavior (e.g., a person's step direction and step distance when a robot intrudes upon their personal space). Joosse et al. (2021) used this instrument to measure people's responses to and tolerance of personal space invasion when being approached by agents at varying speeds.

4.4 Sociability

We define sociability as a robot's conformity to complex, often nuanced, social conventions in its navigational behavior. Previously, we have described various metrics used to measure motion-level social conventions, such as approach velocity, approach pose, invasion of personal space, or passing on the dominant side (e.g., Truong and Ngo, 2016; Guzzi et al., 2013b; Yang and Peters, 2019; Pacchierotti et al., 2006). However, there exist more complex social norms around navigation-based interactions, such as elevator etiquette, waiting in a queue, asking permission to pass, and observing right-of-way at four-way intersections. A robot may move in a natural and appropriate manner that does not cause discomfort, but still violates expected, high-level social norms. For example, a robot may enter an elevator full of people in a perfectly smooth and natural fashion without first letting anyone inside leave; while the robot does not exhibit any unnaturalness or cause discomfort by violating motion-level social conventions, it breaks higher-level social norms that most people expect when riding an elevator. Measuring these high-level social norms would allow for a more holistic understanding of the impact of robot presence on humans; however, measuring sociability remains largely difficult and is considered one of the key challenges in the field of socially-aware navigation (Mavrogiannis et al., 2021).

The Perceived Social Intelligence (PSI) scales proposed by Barchard et al. (2018); Barchard et al. (2020) evaluate 20 aspects of robotic social intelligence. For instance, the Social Competence (SOC) scale consists of four items: 1) social competence, 2) social awareness, 3) social insensitivity (reversed), and 4) strong social skills. PSI scales have been used in previous evaluations of socially-aware navigation (e.g., Barchard et al., 2020); recently, Banisetty and Williams (2021) used the perceived safety scale from the Godspeed questionnaire in conjunction with PSI to evaluate how a robot's spatial motions may communicate social norms during a pandemic via an online study. Additionally, it has been determined that robots using socially-aware navigation planners are perceived to be more socially intelligent as measured by PSI than those using traditional navigation planners (Honour et al., 2021).

In addition to using validated scales, prior research has employed custom questions relevant to specific evaluation contexts to gauge people's perceptions of robot sociability. For example, Vega et al. (2019a) used three questions—*Is the robot's behavior socially appropriate?*; *Is the robot's behavior friendly?*; and *Does the robot understand the social context and the interaction?*—to evaluate how a mobile robot may interact with people to ask for permission to pass when they block its path. All in all, how best to measure sociability remains unresolved, as opposed to the consensus on metrics for evaluating navigation performance and trajectory similarity.

5 DISCUSSION

In this paper, we review the evaluation protocols—focusing on evaluation methods, scenarios, datasets, and metrics—most commonly used in socially-aware robot navigation with the goal of facilitating further progress in this field, which currently lacks principled frameworks for development and evaluation. Prevalent evaluation methods include simulation experiments followed by experimental demonstration, as well as laboratory and field studies. Controlled experiments, either in simulation or in the physical world, typically focus on a set of primitive scenarios such as passing, crossing, and approaching. Datasets of human movements and trajectories are regularly utilized in developing and evaluating socially-aware navigation policies. Prior works have also explored a range of objective, subjective, and behavioral measures to evaluate navigation performance, naturalness of movement, physical and psychological safety, and sociability. Below, we discuss limitations of the existing evaluation protocols and open problems to solve in future research.

5.1 Limitations of Existing Evaluation Protocols

5.1.1 Evaluation Methods, Scenarios, and Datasets

Recent works on socially-aware navigation rely heavily on datasets and simulation experiments for evaluation (Mavrogiannis et al., 2021); this trend has been accelerated by advances in reinforcement learning and data-driven approaches in general (e.g., Luber et al., 2012; Zhou et al., 2012; Alahi et al., 2014; Alahi et al., 2016; Kretzschmar et al., 2016; Park et al., 2016). However, this type of evaluation makes strong assumptions about human and robot behaviors. For example, in simulation experiments, researchers typically rely on pedestrian behavior models such as Optimal Reciprocal Collision Avoidance (ORCA) (Van Den Berg et al., 2011) (e.g., Chen et al., 2017b; Daza et al., 2021) and the Social Force Model (SFM) (Helbing and Molnár, 1995) (e.g., Katyal et al., 2021). Reciprocal behavior models such as ORCA impose the assumption that each agent is fully aware of its surroundings and the position and velocity of the other agents; this assumption of omniscience does not hold true for a real robot or person (Fraichard and Levesy, 2020). Moreover, agents trained using ORCA and SFM behave much differently than real-life agents (Mavrogiannis et al., 2021) and there exist a multitude of SFM variations (e.g., Moussaïd et al., 2009; Anvari et al., 2015; Truong and Ngo, 2017; Huang et al., 2018; Yang and Peters, 2019); therefore, it is important to ensure comparable settings for training and evaluation when comparing algorithms in simulation experiments.

To add to this concern of agent behavior assumptions, the simulators used in virtual social navigation experiments have their own limitations. While 2D simulators such as Stage (Gerkey et al., 2003) and CrowdNav (Chen C. et al., 2019) are lightweight and easy to extend, they oversimplify and abstract, rendering their results difficult to apply to the real world. Recently, several high-fidelity, photorealistic simulation environments were developed for indoor navigation, such as Matterport 3D (Chang et al., 2017) and Gibson (Xia et al., 2018). These environments offer improved simulations closer to real-world settings; however, generating realistic, grounded human social behaviors in high-fidelity simulation environments is still challenging.

Simulation experiments typically leverage datasets and metrics that quantify performance and similarity as described in **Section 4.2.1**. This reliance on datasets and quantitative metrics assumes that the human behaviors recorded in those datasets represent the optimal behaviors for a robot—despite robots possessing dynamics and dimensions largely dissimilar to humans; at best, it is highly debatable whether an exact copy of human trajectories is socially acceptable for all robots. Finally, as described in **Section 3.1**, simulation experiments are commonly followed by demonstrations with physical robots in a real-world setting; while appropriate for proofs-of-concept, these demonstrations are mainly illustrative and lack statistical rigorousness.

In contrast, laboratory studies allow for controlled experiments with statistical precision. However, such experiments are often simplistic and designed for specific navigational interactions (**Table 2**) in certain settings (e.g., passing interactions in a hallway). Moreover, it is important to note that interaction scenarios are usually evaluated out of context. Take the crossing scenario as an example; although crossing is largely evaluated in an open setting (e.g., circle crossing), people may exhibit very different crossing behaviors in real life, as shaped by their individual objectives, other pedestrians, and the environment (e.g., in an open square or an art gallery). Furthermore, laboratory studies typically rely on convenience sampling for participant recruitment (e.g., college students and local residents), resulting in findings that may have limited generalization to a broader population.

Field studies are arguably the most challenging evaluation method to execute; they require robots to operate robustly and safely in unstructured human environments and naturally involve emergent, unprescribed human-robot interactions. While challenging and costly, field studies can provide rich, and sometimes unexpected, insights that simulation and laboratory studies cannot offer (**Section 3.1.4**).

Going forward, we predict an increased need for bridging algorithmic innovations in simulation and autonomous, real-world interactions. Deploying robots for human interaction, either in the field or in laboratory settings, will help us better understand the true limitations of robotics technology and how people experience and interact with it. We strongly advocate for more laboratory and field studies to productively advance socially-aware robot navigation and develop useful, functional mobile robots.

5.1.2 Evaluation Metrics

Navigation Performance

Socially-aware robot navigation shares many performance metrics with general robot navigation. Conventional performance metrics, such as efficiency and success rate, are commonly reported in the literature of socially-aware robot navigation. For example, path efficiency is the ratio of the optimal path's length to that of the actual path and is used to measure path disturbance to agents (either the robot or human pedestrians), while success rate measures an agent's ability to reach its goal. Though not typically used in evaluating socially-aware navigation, we believe metrics that account for both path efficiency and success rate, such as Success weighted by Path Length (SPL) (Anderson et al., 2018), Success weighted by Number of Actions (SNA) (Chen et al., 2021), and Success weighted by Completion Time (SCT) (Yokoyama et al., 2021), are useful metrics to compare navigation policies. However, these metrics should only be used for comparisons in the same setting, as different settings have different optimalities. All in all, these metrics attempt to sum up navigation trials into singular values; while such abstraction is useful for systematic comparison, it makes the assessment of fine-grained trajectory quality more difficult. To answer questions like what caused a particular defect in efficiency, researchers typically visualize trajectories for more qualitative analysis. However, it is worth noting that the most socially acceptable navigational behaviors are not necessarily efficiency- or performance-oriented.

Naturalness

A common method of measuring naturalness is quantifying the similarity between the robot's or the predicted trajectory and those observed in human data. Average Displacement Error (ADE) and Final Displacement Error (FDE) are conventional metrics for quantifying trajectory differences. Variations of displacement- or distance-based metrics may be employed to highlight certain aspects of navigation; for instance, ADE over the nonlinear portions of a trajectory may capture the effects of navigational interactions (e.g., passing and crossing). These types of metrics are typically used in benchmarking navigation algorithms against provided datasets in simulation experiments. While allowing for reproducible and systematic development and evaluation, this dataset-oriented evaluation protocol has several limitations. First, human navigational behaviors and trajectories are context-dependent. The recorded human behaviors in a dataset are specific to the scenario in which the data was collected; moreover, most datasets only include a limited number of scenarios. Therefore, the generalizability of the evaluated algorithms to different contexts is not adequately captured by these metrics. Second, robots and humans afford distinct navigational behaviors and expectations. At the physical level, robots are quite dissimilar to humans and therefore afford different navigational behaviors, such as moving speed. At the social level, it has been revealed that people exhibit different social expectations toward robots than humans; for instance, empirical data suggests that people are willing to let robots get closer to them than they let fellow humans (Joosse et al., 2021). Finally, the majority of existing datasets are limited to 2D trajectories and neglect the fact that navigational behaviors are multimodal in nature. Such limitations necessitate the inclusion of additional metrics to cover aspects of naturalness like sociability and interaction quality.

Instead of using recorded human trajectories as a gold standard for assessing naturalness, several context-independent metrics have been utilized to measure movement smoothness, which is regarded as an important indicator of naturalness. These metrics usually consider velocity and acceleration profiles and path irregularity, which captures the number of unnecessary turns in a path. However, appropriate interpretation of the results from these metrics requires reference points (e.g., is a path irregularity value of 0.72 "good?") that are difficult to obtain and may depend on various factors such as environmental context and culture.

Discomfort

Discomfort is another key dimension in which socially-aware robot navigation is evaluated; it can be characterized generally by physical and psychological safety. To approximate discomfort, prior works have relied upon spatial models including Hall (1966) theory on proxemics and personal space, *f-formation* for groups (Kendon, 2010), the Social Force Model (SFM) (Helbing and Molnár, 1995), and the Extended Social Force Model (ESFM) (Moussaïd et al., 2009). These models are particularly relevant to and useful in evaluating mobile navigation and spatial relationships; specifically, they have been adapted to define safety zones and identify abnormal behaviors (e.g., invading personal space) that may cause discomfort. For instance, prior research has used the Social Individual Index (SII), a numerical metric derived from spatial models, along with empirically determined thresholds to gauge psychological safety (Truong and Ngo, 2017). However, spatial model-based metrics are limited in several ways. First, all agents are assumed to be identical (e.g., possessing the same personal space and social forces), neglecting individual differences observed in the real world; for instance, how people distance themselves from others depends upon personal relationships, individual characteristics, interaction contexts, and cultural norms. Second, common spatial models do not have sufficient granularity to represent environmental contexts. As an example, in SFM, repulsive forces from the environment are all treated the same; however, people move and interact differently in different contexts, and are therefore likely to have varying levels of discomfort tolerance in response to robot navigational behaviors. Third, it is difficult to encode high-level social norms (e.g., sociability) into these spatial models. Altogether, spacial model-based metrics are limited in their ability to represent, simulate, and quantify complex, nuanced social behaviors that humans expect and exhibit in navigation.

In addition to using the aforementioned metrics, discomfort may be measured by self-report ratings [e.g., the perceived safety subscale from the Godspeed questionnaire (Bartneck et al., 2008)] and behavioral indices [e.g., the BEHAVE-II instrument (Joosse et al., 2013)]. These measures are effective in revealing people's subjective experiences and genuine behavioral responses, which may not be accurately represented by objective metrics derived from spatial models. It is worth noting that these subjective and behavioral measures are collected after experiment completion and are consequently unsuitable for learning or adapting robot behavior in real time; however, some of the behavioral measures (e.g., step distance, facial expressions, and eye gaze) from BEHAVE-II may be calculated using computer vision techniques and therefore have the potential to be utilized in real-time behavioral adaptation.

Sociability

Sociability is a complex construct that characterizes a robot's conformity to high-level social conventions, which are conditioned on varying factors such as culture, interaction and environmental contexts, and individual characteristics (e.g., gender); as a result, there are no predetermined sets of high-level social conventions. Therefore, research thus far has explored social conventions that are by and large cherry-picked by the researchers themselves. For example, Pacchierotti et al. (2006) defined a set of social rules for hallway interactions, suggesting that a robot should 1) signal its intention by proactively moving to the right; 2) stay as far away from humans as the width of the hallway allows; and 3) wait until a person completely passes by before resuming normal navigation in order to avoid causing discomfort. Salek Shahrezaie et al. (2021) emphasized that social rules differ based on environmental contexts; for instance, a robot will need to behave differently in galleries, hallways, and around vending machines. The wide range of influencing factors on sociability makes it challenging to adopt a uniform evaluation standard or set of metrics. As a consequence, most prior works adopted an ad hoc approach, using custom questions to assess sociability (e.g., Vega et al., 2019a). More recently, Perceived Social Intelligence (PSI) scales (Barchard et al., 2020) offer an initial point for benchmarking the subjective construct of sociability. In order to productively advance socially-aware navigation, however, further research is required to develop comprehensive instruments specifically designed to measure sociability and higher-level social skills in the context of navigational interactions.

5.2 Open Problems and Opportunities

5.2.1 Diverse, Dynamic Human Models and Long-Term Effects

As discussed in **Section 5.1.1**, there are several limitations to simulation-based evaluation, the most notable of which being homogeneity—all agents are driven by a static behavior engine—and omniscience—all agents have full awareness of their surroundings (Fraichard and Levesy, 2020); these assumptions are a result of the oversimplification and abstraction built into simulators. Moreover, most spatial models for crowd behavior and proxemics are derived from population data; consequently, the experiments and simulations using them often do not support a sufficiently diverse representation of different groups of people (Hurtado et al., 2021). Indeed, humans are naturally diverse and their behaviors and expectations change over time and according to complex factors like individual traits, cultures, and contexts. For example, abundant empirical evidence has demonstrated how age (e.g., Nomura et al., 2009; Flandorfer, 2012), personality (e.g., Walters et al., 2005; Robert, 2018), gender (e.g., Flandorfer, 2012; Strait et al., 2015), and cultural (e.g., Lim et al., 2020) differences may affect people's perceptions of and interactions with robots. Moreover, similar to how people gradually change their behaviors (e.g., standing closer when talking to each other) to reflect developments in a relationship (Altman and Taylor, 1973), robots must also evolve their behaviors—as opposed to exhibiting behaviors uniformly over time—to match their relationships and promote rapport with users. Not only must we develop behavior models to account for gradual changes in relationships, but we must conduct more longitudinal studies to explore how people's experiences with, perceptions of, and behaviors toward robots change over long periods of time. Buchner et al. (2013) demonstrated that a person's experience with a collaborative robot clearly changes over the course of a year; will we see similar effects in navigational human-robot interactions? Ultimately, we have three recommendations for future research:

- Enrich pedestrian models: Although there are limitations to simulation-based approaches to socially-aware navigation, these approaches allow for rapid development and systematic benchmarking and are particularly useful for early-stage validation. However, future simulation-based research must augment pedestrian models to account for human diversity; this may be achieved by including variables to represent the influencing factors we previously discussed and by introducing parameters to regulate said variables over time and according to interaction contexts.
- Examine longitudinal effects: Our understanding of the longitudinal effects of navigational human-robot interactions is fairly limited, yet such knowledge is

critical in developing and integrating mobile robots into real-life environments with the goal of interacting with and assisting people in their daily lives. As the field of socially-aware robot navigation continues to evolve, research efforts should increasingly concentrate on conducting longitudinal field studies.

- Measure and report individual characteristics: As previously mentioned, many characteristics and factors demonstrably influence general human-robot interaction. To collectively advance our understanding of navigational human-robot interaction, we encourage future works to collect and report data on individual characteristics (e.g., age, personality, gender, and culture) and how they relate to the metrics of socially-aware navigation.

5.2.2 Evaluating Mobile Robots of Different Forms

In this paper, we focus on the evaluation of socially-aware navigation in typical mobile robots that move around and interact with people in human environments, such as indoor or outdoor delivery robots. However, mobile robots can take many forms, interactions with humans can happen in different settings (e.g., where people are "on" or "inside" the robot), and human environments can include larger-scale infrastructures such as roads and highways. In particular, our review does not address two notable classes of "robot": robotic wheelchairs and autonomous vehicles. While these two categories share various characteristics in terms of socially-aware navigation, they necessitate additional evaluation considerations and methods.

Similar to traditional mobile robots, robotic wheelchairs must consider the people around them when moving through human environments (e.g., Kretzschmar et al., 2016); as such, various evaluation considerations and metrics discussed in this paper may be adapted for this category of "robot." However, robotic wheelchairs must also take into account additional considerations for their direct users; for instance, Morales et al. (2015) explored ways of including human factors (e.g., user visibility of the environment) when planning paths for a robotic wheelchair and evaluated how comfortable users felt during the ride. In support of greater accessibility and equity, more research is needed to investigate developing and evaluating methods that enable people who are robotic wheelchair-bound to engage in social interactions with individuals or groups of people (e.g., joining or following a social group) (e.g., Escobedo et al., 2014); as such, robotic wheelchairs should consider both users' and surrounding pedestrians' social signals (e.g., intent to interact). The navigation evaluation should also include behavioral indices that capture such nuanced social dynamics. Moreover, as robotic wheelchair users have varying physical disabilities, the development and evaluation of socially-aware navigation capabilities for robotic wheelchairs must pay closer attention to individual needs. Accordingly, custom metrics may be more appropriate for evaluation, as opposed to relying upon a rigid set of standardized evaluation protocols. Detailed reporting of user characteristics and specific needs would help contextualize evaluation results.

Autonomous vehicles (AVs) are up-and-coming "mobile robots" that interact with humans, including the "driver," pedestrians, and other motorists on the road. Like traditional delivery robots, AVs must drive in a safe and predictable manner, but beyond excellent safety protocols and autonomous capabilities, AVs also require critical social awareness; social interactions underlie all pedestrian-vehicle interactions (Rasouli and Tsotsos, 2020) and even AV-AV interactions are considered social coordination events (Schwarting et al., 2019). Similar to evaluating robotic wheelchair applications, the evaluation of AV technology must consider a range of stakeholders, including pedestrians (e.g., Randhavane et al., 2019; Camara et al., 2021), bicyclists (e.g., Rahman et al., 2021), and other drivers (e.g., Schwarting et al., 2019). However, AV evaluation poses additional challenges (e.g., legal regulation for high-stake, life-critical applications) and has different considerations and norms (e.g., following traffic rules). To mitigate safety concerns, recent research has leveraged modern immersive technology such as virtual reality (VR) (e.g., Goedicke et al., 2018; Mahadevan et al., 2019; Camara et al., 2021) when evaluating socially-aware AVs; for instance, Camara et al. (2021) did their user study in a virtual reality setting to evaluate pedestrians' behavior when crossing road with vehicles present. Similar to the evaluation for mobile robots, it is very important to measure the subjective perception of pedestrian-vehicle interactions (Mahadevan et al., 2019) and consider unique spatial interactions in AV applications.

To conclude, we expect to see more autonomous mobile technologies coexisting with people in their daily lives. While these technologies—ranging from mobile service robots and robotic wheelchairs to autonomous vehicles—may have domain-specific considerations for their development and evaluation, social awareness will be vital to the successful adoption of these technologies by the general population.

6 CONCLUSION

As the field of socially-aware navigation continues to evolve, it is vital to cultivate principled frameworks for the development and evaluation of mobile robots that aim to navigate in human environments in an efficient, safe, and socially acceptable manner. In this paper, we review the evaluation protocols commonly used in socially-aware robot navigation as an effort toward developing a principled evaluation framework. Our review highlights the advantages and disadvantages of different evaluation methods and metrics; in particular, while simulation experiments allow for agile development and systematic comparisons, laboratory and field studies can offer valuable insights into navigational human-robot interactions. Moreover, objective, subjective, and behavioral metrics used together offer a more comprehensive view of robot navigation performance and user experience than individual sets of metrics alone. By reviewing evaluation protocols for socially-aware robot navigation, this paper contributes to the broader vision of successful integration of socially-aware mobile technologies into our daily lives.

AUTHOR CONTRIBUTIONS

YG and C-MH co-wrote this manuscript; both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Johns Hopkins University Institute for Assured Autonomy.

ACKNOWLEDGMENTS

We would like to thank Jaimie Patterson for proofreading this paper.

REFERENCES

Ahmadi, E., Meghdari, A., and Alemi, M. (2020). A Socially Aware SLAM Technique Augmented by Person Tracking Module. *J. Intell. Robot. Syst.* 99, 3–12. doi:10.1007/s10846-019-01120-z

Aiello, J. R. (1977). A Further Look at Equilibrium Theory: Visual Interaction as a Function of Interpersonal Distance. *J. Nonverbal Behav.* 1, 122–140. doi:10.1007/bf01145461

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., and Savarese, S. (2016). "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas: IEEE) 2016-Decem, 961–971. doi:10.1109/cvpr.2016.110

Alahi, A., Ramanathan, V., and Fei-Fei, L. (2014). Socially-aware Large-Scale Crowd Forecasting. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2211–2218. doi:10.1109/cvpr.2014.283

Althaus, P., Ishiguro, H., Kanda, T., Miyashita, T., and Christensen, H. I. (2004). "Navigation for Human-Robot Interaction Tasks," in IEEE International Conference on Robotics and Automation, 2004. Proceeding, New Orleans, LA, USA, 26 April-1 May 2004, 1894–1900. doi:10.1109/ROBOT.2004.1308100

Altman, I., and Taylor, D. A. (1973). *Social Penetration: The Development of Interpersonal Relationships*. New York: Holt Rinehart & Winston.

Amaoka, T., Laga, H., and Nakajima, M. (2009). "Modeling the Personal Space of Virtual Agents for Behavior Simulation," in 2009 International Conference on CyberWorlds 7-11 Sept. 2009,Bradford, UK 364–370. doi:10.1109/cw.2009.19

Amirian, J., Hayet, J.-B., and Pettre, J. (2019). "Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories with GANs," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (Long Beach: IEEE), 2964–2972. doi:10.1109/CVPRW.2019.00359

Anderson, C., Du, X., Vasudevan, R., and Johnson-Roberson, M. (2019). "Stochastic Sampling Simulation for Pedestrian Trajectory Prediction," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Macau, China 3-8 Nov. 2019, 1–8. doi:10.1109/iros40897.2019.8967857

Anderson, P., Chang, A., Chaplot, D. S., Dosovitskiy, A., Gupta, S., Koltun, V., et al. (2018). On Evaluation of Embodied Navigation Agents. ArXiv180706757 Cs.

Anvari, B., Bell, M. G. H., Sivakumar, A., and Ochieng, W. Y. (2015). Modelling Shared Space Users via Rule-Based Social Force Model. *Transportation Res. C: Emerging Tech.* 51, 83–103. doi:10.1016/j.trc.2014.10.012

Aria, M., and Cuccurullo, C. (2017). Bibliometrix : An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetrics* 11, 959–975. doi:10.1016/j.joi.2017.08.007

Aroor, A., Epstein, S. L., and Korpan, R. (2017). "MengeROS: A Crowd Simulation Tool for Autonomous Robot Navigation," in 2017 AAAI Fall Symposium Series, 3.

Avelino, J., Garcia-Marques, L., Ventura, R., and Bernardino, A. (2021). Break the Ice: A Survey on Socially Aware Engagement for Human-Robot First Encounters. *Int. J. Soc. Robotics* 13, 1851–1877. doi:10.1007/s12369-020-00720-2

Bachiller, P., Rodriguez-Criado, D., Jorvekar, R. R., Bustos, P., Faria, D. R., and Manso, L. J. (2021). A Graph Neural Network to Model Disruption in Human-Aware Robot Navigation. *Multimed. TOOLS Appl*, 1–19. doi:10.1007/s11042-021-11113-6

Baldassare, M. (1978). Human Spatial Behavior. *Annu. Rev. Sociol.* 4, 29–56. doi:10.2307/294596410.1146/annurev.so.04.080178.000333

Bandini, S., Gorrini, A., and Vizzari, G. (2014). Towards an Integrated Approach to Crowd Analysis and Crowd Synthesis: A Case Study and First Results. *Pattern Recognition Lett.* 44, 16–29. doi:10.1016/j.patrec.2013.10.003

Banisetty, S. B., Forer, S., Yliniemi, L., Nicolescu, M., and Feil-Seifer, D. (2021). Socially-aware Navigation: A Non-linear Multi-Objective Optimization Approach. ArXiv191104037 Cs. doi:10.1145/3453445

Banisetty, S. B., and Williams, T. (2021). "Implicit Communication through Social Distancing: Can Social Navigation Communicate Social Norms," in Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 499–504.

[Dataset] Barchard, K. A., Lapping-Carr, L., Shane, R. W., Banisetty, S. B., and Feil-Seifer, D. (2018). Perceived Social Intelligence (PSI) Scales Test Manual.

Barchard, K. A., Lapping-Carr, L., Westfall, R. S., Fink-Armold, A., Banisetty, S. B., and Feil-Seifer, D. (2020). Measuring the Perceived Social Intelligence of Robots. *J. Hum.-Robot Interact.* 9, 1–29. doi:10.1145/3415139

Bartneck, C., Kulic, D., and Croft, E. (2008). Measuring the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Tech. Rep.* 8.

Bastani, V., Campo, D., Marcenaro, L., and Regazzoni, C. (2015). "Online Pedestrian Group Walking Event Detection Using Spectral Analysis of Motion Similarity Graph," in 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1–5. doi:10.1109/avss.2015.7301744

Batista, M. R., Macharet, D. G., and Romero, R. A. F. (2020). Socially Acceptable Navigation of People with Multi-Robot Teams. *J. Intell. Robot. Syst.* 98, 481–510. doi:10.1007/s10846-019-01080-4

Benfold, B., and Reid, I. (2011). "Stable Multi-Target Tracking in Real-Time Surveillance Video," in CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011, 3457–3464. doi:10.1109/CVPR.2011.5995667

Bera, A., Randhavane, T., and Manocha, D. (2019). "The Emotionally Intelligent Robot:improving Socially-Aware Human Prediction in Crowded Environments," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Bera, A., Randhavane, T., Prinja, R., and Manocha, D. (2017). "SocioSense: Robot Navigation Amongst Pedestrians with Social and Psychological Constraints," in IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24-28 Sept. 2017, 7018–7025. doi:10.1109/iros.2017.8206628

Bisagno, N., Zhang, B., and Conci, N. (2019). "Group LSTM: Group Trajectory Prediction in Crowded Scenarios," in Lecture Notes in Computer Science,Computer Vision – ECCV 2018 Workshops, 213–225. doi:10.1007/978-3-030-11015-4_18

Biswas, A., Wang, A., Silvera, G., Steinfeld, A., and Admoni, H. (2021). SocNavBench: A Grounded Simulation Testing Framework for Evaluating Social Navigation. ArXiv Prepr. ArXiv210300047.

Boldrer, M., Palopoli, L., and Fontanelli, D. (2020). "Socially-aware Multi-Agent Velocity Obstacle Based Navigation for Nonholonomic Vehicles," in 2020 Ieee 44th Annual Computers, Software, and Applications Conference (Compsac 2020), 18–25. doi:10.1109/COMPSAC48688.2020.00012

Bolei Zhou, B., Xiaogang Wang, X., and Xiaoou Tang, X. (2012). "Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2871–2878. doi:10.1109/cvpr.2012.6248013

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., et al. (2016). OpenAI Gym. ArXiv160601540 Cs.

Bršćić, D., Kidokoro, H., Suehiro, Y., and Kanda, T. (2015). "Escaping from Children's Abuse of Social Robots," in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Portland: ACM), 59–66. doi:10.1145/2696454.2696468

Buchegger, K., Todoran, G., and Bader, M. (2019). "Safe and Efficient Autonomous Navigation in the Presence of Humans at Control Level," in *Advances in Service and Industrial Robotics*. (Cham: Springer), 67, 504–511. RAAD 2018. doi:10.1007/978-3-030-00232-9_53

Buchner, R., Wurhofer, D., Weiss, A., and Tscheligi, M. (2013). "Robots in Time: How User Experience in Human-Robot Interaction Changes over Time," in International Conference on Social Robotics, 138–147. doi:10.1007/978-3-319-02675-6_14

Burgard, W., Cremers, A. B., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., et al. (1998). "The Interactive Museum Tour-Guide Robot," in Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference, AAAI 98, IAAI 98, Madison, Wisconsin, USA, July 26-30, 1998, 11–18.

Butler, J. T., and Agah, A. (2001). Psychological Effects of Behavior Patterns of a mobile Personal Robot. *Auton. Robots* 10, 185–202. doi:10.1023/A:1008986004181

Camara, F., Dickinson, P., and Fox, C. (2021). Evaluating Pedestrian Interaction Preferences with a Game Theoretic Autonomous Vehicle in Virtual Reality. *Transportation Res. F: Traffic Psychol. Behav.* 78, 410–423. doi:10.1016/j.trf.2021.02.017

Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J. (2017). "The Robotic Social Attributes Scale (RoSAS): Development and Validation," in 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 254–262.

Chadalavada, R. T., Andreasson, H., Schindler, M., Palm, R., and Lilienthal, A. J. (2020). Bi-directional Navigation Intent Communication Using Spatial Augmented Reality and Eye-Tracking Glasses for Improved Safety in Human-Robot Interaction. *Robotics and Computer-Integrated Manufacturing* 61, 101830. doi:10.1016/j.rcim.2019.101830

Chang, A., Dai, A., Funkhouser, T., Halber, M., Nießner, M., Savva, M., et al. (2017). Matterport3D: Learning from RGB-D Data in Indoor Environments. ArXiv170906158 Cs. doi:10.1109/3dv.2017.00081

Charalampous, K., Kostavelis, I., and Gasteratos, A. (2017). Recent Trends in Social Aware Robot Navigation: A Survey. *Robotics Autonomous Syst.* 93, 85–104. doi:10.1016/j.robot.2017.03.002

Charalampous, K., Kostavelis, I., and Gasteratos, A. (2016). Robot Navigation in Large-Scale Social Maps: An Action Recognition Approach. *Expert Syst. Appl.* 66, 261–273. doi:10.1016/j.eswa.2016.09.026

Chavdarova, T., Baque, P., Bouquet, S., Maksai, A., Jose, C., Bagautdinov, T., et al. (2018). WILDTRACK: A Multi-Camera HD Dataset for Dense Unscripted Pedestrian Detection. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 5030–5039. doi:10.1109/cvpr.2018.00528

Chen, C., Hu, S., Nikdel, P., Mori, G., and Savva, M. (2020). "Relational Graph Learning for Crowd Navigation," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Las Vegas: IEEE), 10007–10013. doi:10.1109/IROS45743.2020.9340705

Chen, C., Liu, Y., Kreiss, S., and Alahi, A. (2019). "Crowd-robot Interaction: Crowd-Aware Robot Navigation with Attention-Based Deep Reinforcement Learning," in Proceedings - IEEE International Conference on Robotics and Automation, Montreal, QC, Canada, 20-24 2019-May, 6015–6022. doi:10.1109/icra.2019.8794134

Chen, C., Majumder, S., Al-Halah, Z., Gao, R., Ramakrishnan, S. K., and Grauman, K. (2021). Learning to Set Waypoints for Audio-Visual Navigation. ArXiv200809622 Cs.

Chen, K., de Vicente, J. P., Sepulveda, G., Xia, F., Soto, A., Vazquez, M., et al. (2019). A Behavioral Approach to Visual Navigation with Graph Localization Networks. ArXiv190300445 Cs. doi:10.15607/rss.2019.xv.010

Chen, Y. F., Everett, M., Liu, M., and How, J. P. (2017a). "Socially Aware Motion Planning with Deep Reinforcement Learning," in IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24-28 Sept. 2017, 1343–1350. doi:10.1109/iros.2017.8202312

Chen, Y. F., Liu, M., Everett, M., and How, J. P. (2017b). "Decentralized Non-communicating Multiagent Collision Avoidance with Deep Reinforcement Learning," in Proceedings - IEEE International Conference on Robotics and Automation (Marina Bay Sands, Singapore: IEEE), 285–292. doi:10.1109/icra.2017.7989037

Chuang, T.-K., Wang, H.-C., Lin, N.-C., Chen, J.-S., Hung, C.-H., Huang, Y.-W., et al. (2018). "Deep Trail-Following Robotic Guide Dog in Pedestrian Environments for People Who Are Blind and Visually Impaired - Learning from Virtual and Real Worlds," in 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21-25 May 2018, 5849–5855. doi:10.1109/icra.2018.8460994

Claes, D., and Tuyls, K. (2018). Multi Robot Collision Avoidance in a Shared Workspace. *Auton. Robot* 42, 1749–1770. doi:10.1007/s10514-018-9726-5

Curtis, S., and Manocha, D. (2014). "Pedestrian Simulation Using Geometric Reasoning in Velocity Space," in *Pedestrian and Evacuation Dynamics 2012* (Cham: Springer International Publishing), 875–890. doi:10.1007/978-3-319-02447-9_73

Daza, M., Barrios-Aranibar, D., Diaz-Amado, J., Cardinale, Y., and Vilasboas, J. (2021). An Approach of Social Navigation Based on Proxemics for Crowded Environments of Humans and Robots. *Micromachines* 12, 193. doi:10.3390/mi12020193

Du, Y., Hetherington, N. J., Oon, C. L., Chan, W. P., Quintero, C. P., Croft, E., et al. (2019). "Group Surfing: A Pedestrian-Based Approach to Sidewalk Robot Navigation," in Proceedings - IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20-24 May 2019, 6518–6524. doi:10.1109/ICRA.2019.8793608

Dynnikov, I., and Wiest, B. (2007). On the Complexity of Braids. *J. Eur. Math. Soc.* 9, 801–840. doi:10.4171/jems/98

Efran, M. G., and Cheyne, J. A. (1973). Shared Space: The Co-operative Control of Spatial Areas by Two Interacting Individuals. *Can. J. Behav. Science/Revue canadienne des Sci. du comportement* 5, 201–210. doi:10.1037/h0082345

Escobedo, A., Spalanzani, A., and Laugier, C. (2014). "Using Social Cues to Estimate Possible Destinations when Driving a Robotic Wheelchair," in IEEE International Conference on Intelligent Robots and Systems, 3299–3304. doi:10.1109/iros.2014.6943021

Fang, F., Shi, M., Qian, K., Zhou, B., and Gan, Y. (2020). A Human-Aware Navigation Method for Social Robot Based on Multi-Layer Cost Map. *Int. J. Intell. Robot. Appl.* 4, 308–318. doi:10.1007/s41315-020-00125-4

Federici, M. L., Gorrini, A., Manenti, L., and Vizzari, G. (2012). Data Collection for Modeling and Simulation: Case Study at the university of milan-bicocca. *Lecture Notes Comput. Science,Cellular Automata*, 699–708. doi:10.1007/978-3-642-33350-7-7210.1007/978-3-642-33350-7_72

Fei, C. S., Fai, Y. C., Ming, E. S. L., Yong, L. T., Feng, D., and Hua, P. C. J. (2019). Neural-network Based Adaptive Proxemics-Costmap for Human-Aware Autonomous Robot Navigation. *Int. J. Integr. Eng.* 11, 101–111. doi:10.30880/ijie.2019.11.04.011

Ferrer, G., Garrell, A., and Sanfeliu, A. (2013a). Robot Companion: A Social-Force Based Approach with Human Awareness-Navigation in Crowded Environments. *IEEE Int. Conf. Intell. Robots Syst.*, 1688–1694. doi:10.1109/iros.2013.6696576

Ferrer, G., Garrell, A., and Sanfeliu, A. (2013b). "Social-aware Robot Navigation in Urban Environments," in 2013 European Conference on Mobile Robots, Barcelona, Spain, 25-27 Sept. 2013, 331–336. doi:10.1109/ecmr.2013.6698863

Ferrer, G., Zulueta, A. G., Cotarelo, F. H., and Sanfeliu, A. (2017). Robot Social-Aware Navigation Framework to Accompany People Walking Side-By-Side. *Auton. Robot* 41, 775–793. doi:10.1007/s10514-016-9584-y

Flandorfer, P. (2012). Population Ageing and Socially Assistive Robots for Elderly Persons: The Importance of Sociodemographic Factors for User Acceptance. *Int. J. Popul. Res.* 2012, 1–13. doi:10.1155/2012/829835

Forer, S., Banisetty, S. B., Yliniemi, L., Nicolescu, M., and Feil-Seifer, D. (2018). Socially-aware Navigation Using Non-linear Multi-Objective Optimization. *IEEE Int. Conf. Intell. Robots Syst.* 1, 8126–8133. doi:10.1109/iros.2018.8593825

Fraichard, T. (2007). "A Short Report about Motion Safety," in IEEE Int. Conf. on Robotics and Automation, 10–14.

Fraichard, T., and Levesy, V. (2020). From Crowd Simulation to Robot Navigation in Crowds. *IEEE Robot. Autom. Lett.* 5, 729–735. doi:10.1109/LRA.2020.2965032

Fuse, Y., and Tokumaru, M. (2020). Navigation Model for a Robot as a Human Group Member to Adapt to Changing Conditions of Personal Space. *JACIII* 24, 621–629. doi:10.20965/jaciii.2020.p0621

Gaydashenko, A., Kudenko, D., and Shpilman, A. (2018). "A Comparative Evaluation of Machine Learning Methods for Robot Navigation through Human Crowds," in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 553–557. doi:10.1109/ICMLA.2018.00089

Gérin-Lajoie, M., Richards, C. L., Fung, J., and McFadyen, B. J. (2008). Characteristics of Personal Space during Obstacle Circumvention in Physical and Virtual Environments. *Gait & Posture* 27, 239–247. doi:10.1016/j.gaitpost.2007.03.015

Gerkey, B. P., Vaughan, R. T., and Howard, A. (2003). *The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems*. Coimbra: IEEE, 7.

Goedicke, D., Li, J., Evers, V., and Ju, W. (2018). "VR-OOM," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal: ACM), 1–11. doi:10.1145/3173574.3173739

Gonon, D. J., Paez-Granados, D., and Billard, A. (2021). Reactive Navigation in Crowds for Non-holonomic Robots with Convex Bounding Shape. *IEEE Robot. Autom. Lett.* 6, 4728–4735. doi:10.1109/LRA.2021.3068660

Grant, M. J., and Booth, A. (2009). A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies. *Health Inf. Libr. J.* 26, 91–108. doi:10.1111/j.1471-1842.2009.00848.x

Greenberg, C. I., Strube, M. J., and Myers, R. A. (1980). A Multitrait-Multimethod Investigation of Interpersonal Distance. *J. Nonverbal Behav.* 5, 104–114. doi:10.1007/bf00986513

Grzeskowiak, F., Gonon, D., Dugas, D., Paez-Granados, D., Chung, J., Nieto, J., et al. (2021). Crowd against the Machine: A Simulation-Based Benchmark Tool to Evaluate and Compare Robot Capabilities to Navigate a Human Crowd. ArXiv210414177 Cs. doi:10.1109/icra48506.2021.9561694

Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., and Alahi, A. (2018). "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2255–2264. doi:10.1109/cvpr.2018.00240

Guzzi, J., Giusti, A., Gambardella, L. M., and Di Caro, G. A. (2013a). "Local Reactive Robot Navigation: A Comparison between Reciprocal Velocity Obstacle Variants and Human-like Behavior," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2622–2629. doi:10.1109/IROS.2013.6696726

Guzzi, J., Giusti, A., Gambardella, L. M., Theraulaz, G., and Di Caro, G. A. (2013b). "Human-friendly Robot Navigation in Dynamic Environments," in 2013 IEEE International Conference on Robotics and Automation (Karlsruhe: IEEE), 423–430. doi:10.1109/icra.2013.6630610

Hacinecipoglu, A., Konukseven, E. I., and Koku, A. B. (2020). Multiple Human Trajectory Prediction and Cooperative Navigation Modeling in Crowded Scenes. *Intel Serv. Robotics* 13, 479–493. doi:10.1007/s11370-020-00333-8

Hall, E. T. (1966). *The Hidden Dimension*. Garden City, NY: Doubleday.

Hawes, N., Burbridge, C., Jovan, F., Kunze, L., Lacerda, B., Mudrova, L., et al. (2017). The STRANDS Project: Long-Term Autonomy in Everyday Environments. *IEEE Robot. Automat. Mag.* 24, 146–156. doi:10.1109/MRA.2016.2636359

Helbing, D., and Molnár, P. (1995). Social Force Model for Pedestrian Dynamics. *Phys. Rev. E* 51, 4282–4286. doi:10.1103/physreve.51.4282

Honig, S. S., Oron-Gilad, T., Zaichyk, H., Sarne-Fleischmann, V., Olatunji, S., and Edan, Y. (2018). Toward Socially Aware Person-Following Robots. *IEEE Trans. Cogn. Dev. Syst.* 10, 936–954. doi:10.1109/tcds.2018.2825641

Honour, A., Banisetty, S. B., and Feil-Seifer, D. (2021). "Perceived Social Intelligence as Evaluation of Socially Navigation," in Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (Boulder: ACM), 519–523. doi:10.1145/3434074.3447226

Huang, C.-M., Iio, T., Satake, S., and Kanda, T. (2014). "Modeling and Controlling Friendliness for an Interactive Museum Robot," in *Robotics: Science and Systems*. Zaragoza: RSS, 12–16. doi:10.15607/rss.2014.x.025

Huang, L., Gong, J., Li, W., Xu, T., Shen, S., Liang, J., et al. (2018). Social Force Model-Based Group Behavior Simulation in Virtual Geographic Environments. *ISPRS Int. J. Geo-inf.* 7. doi:10.3390/ijgi7020079

Hurtado, J. V., Londoño, L., and Valada, A. (2021). From Learning to Relearning: A Framework for Diminishing Bias in Social Robot Navigation. *Front. Robot. AI* 8, 650325. doi:10.3389/frobt.2021.650325

Jin, J., Nguyen, N. M., Sakib, N., Graves, D., Yao, H., and Jagersand, M. (2019). Mapless Navigation Among Dynamics with Social-Safety-Awareness: A Reinforcement Learning Approach from 2D Laser Scans.

Johnson, C., and Kuipers, B. (2018). "Socially-aware Navigation Using Topological Maps and Social Norm Learning," in AIES '18 Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 151–157. doi:10.1145/3278721.3278772

Joosse, M., Lohse, M., Berkel, N. V., Sardar, A., and Evers, V. (2021). Making Appearances. *J. Hum.-Robot Interact.* 10, 1–24. doi:10.1145/3385121

Joosse, M., Sardar, A., Lohse, M., and Evers, V. (2013). BEHAVE-II: The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot. *Int. J. Soc. Robotics* 5, 379–388. doi:10.1007/s12369-013-0191-1

Kamezaki, M., Kobayashi, A., Yokoyama, Y., Yanagawa, H., Shrestha, M., and Sugano, S. (2020). A Preliminary Study of Interactive Navigation Framework with Situation-Adaptive Multimodal Inducement: Pass-By Scenario. *Int. J. Soc. Robotics* 12, 567–588. doi:10.1007/s12369-019-00574-3

Kato, Y., Kanda, T., and Ishiguro, H. (2015). "May I Help You," in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (Portland: Association for Computing Machinery), 35–42. HRI '15. doi:10.1145/2696454.2696463

Katyal, K. D., Popek, K., Hager, G. D., Wang, I.-J., and Huang, C.-M. (2020). "Prediction-based Uncertainty Estimation for Adaptive Crowd Navigation," in International Conference on Human-Computer Interaction, 353–368. doi:10.1007/978-3-030-50334-5_24

Katyal, K., Gao, Y., Markowitz, J., Pohland, S., Rivera, C., Wang, I.-J., et al. (2021). Learning a Group-Aware Policy for Robot Navigation. ArXiv201212291 Cs.

Kendon, A. (2010). "Spacing and Orientation in Co-present Interaction," in *Development of Multimodal Interfaces: Active Listening and Synchrony* (Berlin: Springer), 1–15. doi:10.1007/978-3-642-12397-9_1

Kessler, J., Schroeter, C., and Gross, H.-M. (2011). "Approaching a Person in a Socially Acceptable Manner Using a Fast Marching Planner," in *Intelligent Robotics and Applications, Pt Ii. Vol. 7102 of Lecture Notes in Artificial Intelligence*. Aachen: Springer, 368–377. doi:10.1007/978-3-642-25489-5_36

Khambhaita, H., and Alami, R. (2020). Viewing Robot Navigation in Human Environment as a Cooperative Activity. *Robotics Res.*, 285–300. doi:10.1007/978-3-030-28619-4_25

Kirby, R. (2010). *Social Robot Navigation*. Pittsburgh: ProQuest, 3470165, 232. ProQuest Diss. Theses.

Kivrak, H., Cakmak, F., Kose, H., and Yavuz, S. (2021). Social Navigation Framework for Assistive Robots in Human Inhabited Unknown Environments. *Eng. Sci. Technol. Int. J.* 24, 284–298. doi:10.1016/j.jestch.2020.08.008

Kivrak, H., and Kose, H. (2018). "Social Robot Navigation in Human-Robot Interactive Environments: Social Force Model Approach," in 2018 26th Signal Processing and Communications Applications Conference (SIU) (Cesme-Izmir, Turkey: IEEE), 1–4. doi:10.1109/siu.2018.8404328

Knowles, E. S., et al, B., Haas, S., Hyde, M., and Schuchart, G. E. (1976). Group Size and the Extension of Social Space Boundaries. *J. Personal. Soc. Psychol.* 33, 647–654. doi:10.1037/0022-3514.33.5.647

Kodagoda, S., Sehestedt, S., and Dissanayake, G. (2016). Socially Aware Path Planning for mobile Robots. *ROBOTICA* 34, 513–526. doi:10.1017/S0263574714001611

Kollmitz, M., Hsiao, K., Gaa, J., and Burgard, W. (2015). "Time Dependent Planning on a Layered Social Cost Map for Human-Aware Robot Navigation," in 2015 European Conference on Mobile Robots, ECMR 2015 - Proceedings. doi:10.1109/ecmr.2015.7324184

Kolve, E., Mottaghi, R., Han, W., VanderBilt, E., Weihs, L., Herrasti, A., et al. (2019). AI2-THOR: An Interactive 3D Environment for Visual AI. ArXiv171205474 Cs

Kostavelis, I., Kargakos, A., Giakoumis, D., and Tzovaras, D. (2017). "Robot's Workspace Enhancement with Dynamic Human Presence for Socially-Aware Navigation," in International Conference on Computer Vision Systems, 279–288. doi:10.1007/978-3-319-68345-4_25

Kothari, P., Kreiss, S., and Alahi, A. (2020). *Human Trajectory Forecasting in Crowds: A Deep Learning Perspective*, 1–33.

Kretzschmar, H., Spies, M., Sprunk, C., and Burgard, W. (2016). Socially Compliant mobile Robot Navigation via Inverse Reinforcement Learning. *Int. J. Robotics Res.* 35, 1289–1307. doi:10.1177/0278364915619772

Kruse, T., Basili, P., Glasauer, S., and Kirsch, A. (2012). "Legible Robot Navigation in the Proximity of Moving Humans," in IEEE Workshop on Advanced Robotics and Its Social Impacts. doi:10.1109/arso.2012.6213404

Kruse, T., Pandey, A. K., Alami, R., and Kirsch, A. (2013). Human-aware Robot Navigation: A Survey. *Robotics Autonomous Syst.* 61, 1726–1743. doi:10.1016/j.robot.2013.05.007

Lasota, P. A., Fong, T., and Shah, J. A. (2017). A Survey of Methods for Safe Human-Robot Interaction. *FNT in Robotics* 5, 261–349. doi:10.1561/2300000052

Le, A. V., and Choi, J. (2018). Robust Tracking Occluded Human in Group by Perception Sensors Network System. *J. Intell. Robot Syst.* 90, 349–361. doi:10.1007/s10846-017-0667-6

Lerner, A., Chrysanthou, Y., and Lischinski, D. (2007). Crowds by Example. *Comput. Graphics Forum* 26, 655–664. doi:10.1111/j.1467-8659.2007.01089.x

Li, J., Yang, F., Tomizuka, M., and Choi, C. (2020). *EvolveGraph: Multi-Agent Trajectory Prediction with Dynamic Relational Reasoning*. Neurips, 19.

Li, K., Xu, Y., Wang, J., and Meng, M. Q.-H. (2019). "SARL*: Deep Reinforcement Learning Based Human-Aware Navigation for Mobile Robot in Indoor Environments," in 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), 688–694. doi:10.1109/ROBIO49542.2019.8961764

Liang, J., Patel, U., Sathyamoorthy, A. J., and Manocha, D. (2020). *Realtime Collision Avoidance for mobile Robots in Dense Crowds Using Implicit Multi-Sensor Fusion and Deep Reinforcement Learning*. thaca: arxiv.

Lim, V., Rooksby, M., and Cross, E. S. (2020). Social Robots on a Global Stage: Establishing a Role for Culture during Human–Robot Interaction. *Int. J. Soc. Robot.* 13, 1–27. doi:10.1007/s12369-020-00710-4

Lindner, F. (2016). "How to Count Multiple Personal-Space Intrusions in Social Robot Navigation," in *What Social Robots Can and Should Do* (Amsterdam: IOS PRESS), 323–331. vol. 290 of Frontiers in Artificial Intelligence and Applications. doi:10.3233/978-1-61499-708-5-323

Liu, L., Dugas, D., Cesari, G., Siegwart, R., and Dube, R. (2020). "Robot Navigation in Crowded Environments Using Deep Reinforcement Learning," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Las Vegas: IEEE), 5671–5677. doi:10.1109/IROS45743.2020.9341540

Liu, Y., Yan, Q., and Alahi, A. (2020). Social NCE: Contrastive Learning of Socially-Aware Motion Representations. ArXiv201211717 Cs.

Luber, M., Spinello, L., Silva, J., and Arras, K. O. (2012). "Socially-aware Robot Navigation: A Learning Approach," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (Vilamoura-Algarve, Portugal: IEEE), 902–907. doi:10.1109/iros.2012.6385716

Luo, R. C., and Huang, C. (2016). "Human-aware Motion Planning Based on Search and Sampling Approach," in 2016 IEEE WORKSHOP ON ADVANCED ROBOTICS AND ITS SOCIAL IMPACTS (ARSO) (Shanghai, China: IEEE), 226–231. IEEE Workshop on Advanced Robotics and Its Social Impacts. doi:10.1109/arso.2016.7736286

Mahadevan, K., Sanoubari, E., Somanath, S., Young, J. E., and Sharlin, E. (2019). "AV-pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator," in DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference, 475–486. doi:10.1145/3322276.3322328

Majecka, B. (2009). Statistical Models of Pedestrian Behaviour in the Forum. Masters Thesis Sch. Inform. Univ. Edinb.

Manso, L. J., Jorvekar, R. R., Faria, D. R., Bustos, P., and Bachiller, P. (2019). *Graph Neural Networks for Human-Aware Social Navigation*. Ithaca: arxiv.

Martin-Martin, R., Patel, M., Rezatofighi, H., Shenoi, A., Gwak, J., Frankel, E., et al. (2021). JRDB: A Dataset and Benchmark of Egocentric Robot Visual Perception of Humans in Built Environments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1. doi:10.1109/TPAMI.2021.3070543

Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., et al. (2021). Core Challenges of Social Robot Navigation: A Survey. ArXiv210305668 Cs.

Mavrogiannis, C., Hutchinson, A. M., MacDonald, J., Alves-Oliveira, P., Knepper, R. A., and Knepper, R. A. (2019). "Effects of Distinct Robot Navigation Strategies on Human Behavior in a Crowded Environment," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 11-14 March 2019 (Daegu, South Korea: IEEE), 421–430. doi:10.1109/hri.2019.8673115

Mavrogiannis, C. I., Blukis, V., and Knepper, R. A. (2017). "Socially Competent Navigation Planning by Deep Learning of Multi-Agent Path Topologies," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 6817–6824. doi:10.1109/iros.2017.8206601

Mavrogiannis, C. I., Thomason, W. B., and Knepper, R. A. (2018). "Social Momentum," in ACM/IEEE International Conference on Human-Robot Interaction, 361–369. doi:10.1145/3171221.3171255

Mead, R., Atrash, A., and Matarić, M. J. (2011). "Proxemic Feature Recognition for Interactive Robots: Automating Metrics from the Social Sciences," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin, Heidelberg: Springer), Vol. 7072 LNAI, 52–61. doi:10.1007/978-3-642-25504-5_6

Mehran, R., Oyama, A., and Shah, M. (2009). "Abnormal Crowd Behavior Detection Using Social Force Model," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20-25 June 2009 (Miami Beach: IEEE), 935–942. doi:10.1109/cvpr.2009.5206641

Morales, Y., Watanabe, A., Ferreri, F., Even, J., Ikeda, T., Shinozawa, K., et al. (2015). "Including Human Factors for Planning Comfortable Paths," in 2015 IEEE International Conference on Robotics and Automation (ICRA), 6153–6159. doi:10.1109/icra.2015.7140063

Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., and Theraulaz, G. (2009). Experimental Study of the Behavioural Mechanisms Underlying Self-Organization in Human Crowds. *Proc. R. Soc. B.* 276, 2755–2762. doi:10.1098/rspb.2009.0405

Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., and Theraulaz, G. (2010). The Walking Behaviour of Pedestrian Social Groups and its Impact on Crowd Dynamics. *PLoS ONE* 5, e10047–7. doi:10.1371/journal.pone.0010047

Murphy, R. R., and Schreckenghost, D. (2013). "Survey of Metrics for Human-Robot Interaction," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 197–198. doi:10.1109/HRI.2013.6483569

Neggers, M. M. E., Cuijpers, R. H., and Ruijten, P. A. M. (2018). "Comfortable Passing Distances for Robots," in *Social Robotics* (Cham: Springer International Publishing), 431–440. Lecture Notes in Computer Science. doi:10.1007/978-3-030-05204-1_42

Neggers, M. M. E., Cuijpers, R. H., Ruijten, P. A. M., and IJsselsteijn, W. A. (2021). Determining Shape and Size of Personal Space of a Human when Passed by a Robot. *Int. J. Soc. Robotics*, 1–12. doi:10.1007/s12369-021-00805-6

Ngo, H. Q. T., Le, V. N., Thien, V. D. N., Nguyen, T. P., and Nguyen, H. (2020). Develop the Socially Human-Aware Navigation System Using Dynamic Window Approach and Optimize Cost Function for Autonomous Medical Robot. *Adv. Mech. Eng.* 12, 168781402097943. doi:10.1177/1687814020979430

Nishimura, M., and Yonetani, R. (2020). *L2B: Learning to Balance the Safety-Efficiency Trade-Off in Interactive Crowd-Aware Robot Navigation*. Ithaca: arxiv. doi:10.1109/iros45743.2020.9341519

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2009). Age Differences and Images of Robots. *Is* 10, 374–391. doi:10.1075/is.10.3.05nom

Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.-C., Lee, J. T., et al. (2011). "A Large-Scale Benchmark Dataset for Event Recognition in Surveillance Video," in CVPR 2011 (IEEE), 3153–3160. doi:10.1109/CVPR.2011.5995586

Okal, B., and Arras, K. O. (2016). "Learning Socially Normative Robot Navigation Behaviors with Bayesian Inverse Reinforcement Learning," in 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16-21 May 2016, 2889–2895. doi:10.1109/icra.2016.7487452

Okal, B., and Arras, K. O. (2014). "Towards Group-Level Social Activity Recognition for mobile Robots," in Proc. 3rd IROS'2014 Workshop "Assistance Serv. Robot. Hum. Environ.

[Dataset] Okal, B., and Linder, T. (2013). *Pedestrian Simulator*. Freiburg: Social Robotics Lab, University of Freiburg.

Pacchierotti, E., Christensen, H. I., and Jensfelt, P. (2006). "Embodied Social Interaction for Service Robots in Hallway Environments," in *Field and Service Robotics* (Cham: Springer), 293–304.

Pandey, A. K., and Alami, R. (2010). "A Framework towards a Socially Aware mobile Robot Motion in Human-Centered Dynamic Environment," in IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings, 5855–5860. doi:10.1109/iros.2010.5649688

Pandey, A. (2017). Mobile Robot Navigation and Obstacle Avoidance Techniques: A Review. *Int. Robot. Autom. J.* 2, 1–12. doi:10.15406/iratj.2017.02.00023

Papenmeier, F., Uhrig, M., and Kirsch, A. (2019). Human Understanding of Robot Motion: The Role of Velocity and Orientation. *Int. J. Soc. Robotics* 11, 75–88. doi:10.1007/s12369-018-0493-4

Park, H. S., Hwang, J.-J., Niu, Y., and Shi, J. (2016). "Egocentric Future Localization," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas: IEEE), 4697–4705. doi:10.1109/CVPR.2016.508

Pellegrini, S., Ess, A., Schindler, K., and Van Gool, L. (2009). "You'll Never Walk Alone: Modeling Social Behavior for Multi-Target Tracking," in Proceedings of the IEEE International Conference on Computer Vision, 261–268. doi:10.1109/iccv.2009.5459260

Pérez-D'Arpino, C., Liu, C., Goebel, P., Martín-Martín, R., and Savarese, S. (2020). *Robot Navigation in Constrained Pedestrian Environments Using Reinforcement Learning*. Ithaca.

Qian, K., Ma, X., Dai, X., and Fang, F. (2010a). Robotic Etiquette: Socially Acceptable Navigation of Service Robots with Human Motion Pattern Learning and Prediction. *J. Bionic Eng.* 7, 150–160. doi:10.1016/S1672-6529(09)60199-2

Qian, K., Ma, X., Dai, X., and Fang, F. (2010b). Socially Acceptable Pre-collision Safety Strategies for Human-Compliant Navigation of Service Robots. *Adv. Robotics* 24, 1813–1840. doi:10.1163/016918610X527176

Rahman, M. T., Dey, K., Das, S., and Sherfinski, M. (2021). Sharing the Road with Autonomous Vehicles: A Qualitative Analysis of the Perceptions of Pedestrians and Bicyclists. *Transportation Res. Part F: Traffic Psychol. Behav.* 78, 433–445. doi:10.1016/j.trf.2021.03.008

Rajamohan, V., Scully-Allison, C., Dascalu, S., and Feil-Seifer, D. (2019). "Factors Influencing the Human Preferred Interaction distanceFactors Influencing the Human Preferred Interaction Distance," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication. RO-MAN 2019. doi:10.1109/RO-MAN46459.2019.8956404

Randhavane, T., Bera, A., Kubin, E., Wang, A., Gray, K., and Manocha, D. (2019). "Pedestrian Dominance Modeling for Socially-Aware Robot Navigation," in Proceedings - IEEE International Conference on Robotics and Automation, Montreal, QC, Canada, 20-24 May 2019, 5621–5628. doi:10.1109/icra.2019.8794465

Rasouli, A., and Tsotsos, J. K. (2020). Autonomous Vehicles that Interact with Pedestrians: A Survey of Theory and Practice. *IEEE Trans. Intell. Transport. Syst.* 21, 900–918. doi:10.1109/tits.2019.2901817

Repiso, E., Garrell, A., and Sanfeliu, A. (2020). People's Adaptive Side-By-Side Model Evolved to Accompany Groups of People by Social Robots. *IEEE Robot. Autom. Lett.* 5, 2387–2394. doi:10.1109/LRA.2020.2970676

Rios-martinez, J., Renzaglia, A., Spalanzani, A., Martinelli, A., and Laugier, C. (2013). "Navigating between People: A Stochastic Optimization Approach," in 2012 IEEE International Conference on Robotics and Automation (St Paul: IEEE) 231855, 2880–2885. doi:10.1109/icra.2012.6224934

Rios-Martinez, J., Spalanzani, A., and Laugier, C. (2015). From Proxemics Theory to Socially-Aware Navigation: A Survey. *Int. J. Soc. Robotics* 7, 137–153. doi:10.1007/s12369-014-0251-1

Ristani, E., and Tomasi, C. (2015). Tracking Multiple People Online and in Real Time. *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, 444–459. doi:10.1007/978-3-319-16814-2_29

Robert, L. (2018). "Personality in the Human Robot Interaction Literature: A Review and Brief Critique," in *Personality in the Human Robot Interaction Literature: A Review and Brief Critique, Proceedings of the 24th Americas Conference on Information Systems*. Editor L. P. Robert. New Orleans: AMCIS, 16–18.

Robicquet, A., Sadeghian, A., Alahi, A., and Savarese, S. (2016). Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes. *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma* 9912 LNCS, 549–565. doi:10.1007/978-3-319-46484-8_33

Rudenko, A., Palmieri, L., and Arras, K. O. (2017). "Predictive Planning for a mobile Robot in Human Environments," in Proceedings of the Workshop on AI Planning and Robotics: Challenges and Methods (at ICRA 2017).

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., and Arras, K. O. (2019). *Human Motion Trajectory Prediction: A Survey*. Ithaca: arxiv, 1–33.

Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, S. H., and Savarese, S. (2018a). *SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints*. Long Beach: IEEE.

Sadeghian, A., Legros, F., Voisin, M., Vesel, R., Alahi, A., and Savarese, S. (2018b). CAR-net: Clairvoyant Attentive Recurrent Network. *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma* 11215 LNCS, 162–180. doi:10.1007/978-3-030-01252-6_10

Sakoe, H., and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoust. Speech, Signal. Process.* 26, 43–49. doi:10.1109/tassp.1978.1163055

Salek Shahrezaie, R., Banisetty, S. B., Mohammadi, M., and Feil-Seifer, D. (2021). "Towards Deep Reasoning on Social Rules for Socially Aware Navigation," in Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (Boulder: ACM), 515–518. doi:10.1145/3434074.3447225

Samsani, S. S., and Muhammad, M. S. (2021). Socially Compliant Robot Navigation in Crowded Environment by Human Behavior Resemblance Using Deep Reinforcement Learning. *IEEE Robot. Autom. Lett.* 6, 5223–5230. doi:10.1109/LRA.2021.3071954

Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., and Hagita, N. (2009). "How to Approach Humans," in Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction - HRI '09 (La Jolla: ACM Press), 109. doi:10.1145/1514095.1514117

Scandolo, L., and Fraichard, T. (2011). "An Anthropomorphic Navigation Scheme for Dynamic Scenarios," in Proceedings - IEEE International Conference on Robotics and Automation, 809–814. doi:10.1109/icra.2011.5979772

Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S., and Rus, D. (2019). Social Behavior for Autonomous Vehicles. *Proc. Natl. Acad. Sci. USA* 116, 24972–24978. doi:10.1073/pnas.1820676116

Sebastian, M., Banisetty, S. B., and Feil-Seifer, D. (2017). Socially-aware Navigation Planner Using Models of Human-Human Interaction. doi:10.1109/ROMAN.2017.8172334

Senft, E., Satake, S., and Kanda, T. (2020). "Would You Mind Me if I Pass by You," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (Cambridge, United Kingdom: ACM IEEE International Conference on Human-Robot Interaction), 539–547. doi:10.1145/3319502.3374812

Shiomi, M., Zanlungo, F., Hayashi, K., and Kanda, T. (2014). Towards a Socially Acceptable Collision Avoidance for a mobile Robot Navigating Among Pedestrians Using a Pedestrian Model. *Int. J. Soc. Robotics* 6, 443–455. doi:10.1007/s12369-014-0238-y

Sisbot, E., Alami, R., Simeon, T., Dautenhahn, K., Walters, M., Woods, S., et al. (2005). "Navigation in the Presence of Humans," in 5th IEEE-RAS International Conference on Humanoid Robots, 2005 (IEEE-RAS International Conference on Humanoid Robots), 181–188.

Sisbot, E. A., Marin-Urias, L. F., Alami, R., and Simeon, T. (2007). A Human Aware mobile Robot Motion Planner. *IEEE Trans. Robot.* 23, 874–883. doi:10.1109/TRO.2007.904911

Sochman, J., and Hogg, D. C. (2011). Who Knows Who - Inverting the Social Force Model for Finding Groups. *Proc. IEEE Int. Conf. Comput. Vis.*, 830–837. doi:10.1109/iccvw.2011.6130338

Sprute, D., Tönnies, K., and König, M. (2019). A Study on Different User Interfaces for Teaching Virtual Borders to mobile Robots. *Int. J. Soc. Robotics* 11, 373–388. doi:10.1007/s12369-018-0506-3

Stein, P., Spalanzani, A., Santos, V., and Laugier, C. (2016). Leader Following: A Study on Classification and Selection. *Robotics Autonomous Syst.* 75, 79–95. doi:10.1016/j.robot.2014.09.028

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). "Common Metrics for Human-Robot Interaction," in HRI 2006: Proceedings of the 2006 ACM Conference on Human-Robot Interaction, 33–40. doi:10.1145/1121241.1121249

Strait, M., Briggs, P., and Scheutz, M. (2015). "Gender, More So Than Age, Modulates Positive Perceptions of Language-Based Human-Robot Interactions," in 4th International Symposium on New Frontiers in Human Robot Interaction, 21–22.

Sui, Z., Pu, Z., Yi, J., and Xiong, T. (2019). "Formation Control with Collision Avoidance through Deep Reinforcement Learning," in 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14-19 July 2019, 1–8. doi:10.1109/ijcnn.2019.8851906

Sun, S., Zhao, X., Li, Q., and Tan, M. (2020). Inverse Reinforcement Learning-Based Time-dependent A* Planner for Human-Aware Robot Navigation with Local Vision. *Adv. Robotics* 34, 888–901. doi:10.1080/01691864.2020.1753569

Tai, L., Zhang, J., Liu, M., and Burgard, W. (2018). "Socially Compliant Navigation through Raw Depth Inputs with Generative Adversarial Imitation Learning," in Proceedings - IEEE International Conference on Robotics and Automation, 1111–1117. doi:10.1109/icra.2018.8460968

Talebpour, Z., Navarro, I., and Martinoli, A. (2015). "On-board Human-Aware Navigation for Indoor Resource-Constrained Robots: A Case-Study with the ranger," in 2015 Ieee/Sice International Symposium on System Integration (Sii) (Nagoya, Japan: IEEE), 63–68. doi:10.1109/sii.2015.7404955

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., et al. (1999). *MINERVA: A Second-Generation Museum Tour-Guide Robot*. Detroit: IEEE.

Tomari, R., Kobayashi, Y., and Kuno, Y. (2014). "Analysis of Socially Acceptable Smart Wheelchair Navigation Based on Head Cue Information, Procedia Computer Science," in *Medical and Rehabilitation Robotics and Instrumentation (Mrri2013)*, 198–205. vol. 42 of Procedia Computer Science (Amsterdam: Elsevier). doi:10.1016/j.procs.2014.11.052

Torta, E., Cuijpers, R. H., and Juola, J. F. (2013). Design of a Parametric Model of Personal Space for Robotic Social Navigation. *Int. J. Soc. Robotics* 5, 357–365. doi:10.1007/s12369-013-0188-9

Trautman, P., Ma, J., Murray, R. M., and Krause, A. (2015). Robot Navigation in Dense Human Crowds: Statistical Models and Experimental Studies of Human-Robot Cooperation. *Int. J. Robotics Res.* 34, 335–356. doi:10.1177/0278364914557874

Truong, X.-T., and Ngo, T.-D. (2018). "To Approach Humans?": A Unified Framework for Approaching Pose Prediction and Socially Aware Robot

Navigation. *IEEE Trans. Cogn. Dev. Syst.* 10, 557–572. doi:10.1109/tcds.2017.2751963
Truong, X.-T., and Ngo, T.-D. (2016). Dynamic Social Zone Based mobile Robot Navigation for Human Comfortable Safety in Social Environments. *Int. J. Soc. Robotics* 8, 663–684. doi:10.1007/s12369-016-0352-0
Truong, X.-T., and Ngo, T. D. (2017). Toward Socially Aware Robot Navigation in Dynamic and Crowded Environments: A Proactive Social Motion Model. *IEEE Trans. Automat. Sci. Eng.* 14, 1743–1760. doi:10.1109/tase.2017.2731371
Truong, X.-T., Yoong, V. N., and Ngo, T.-D. (2017). Socially Aware Robot Navigation System in Human Interactive Environments. *Intel Serv. Robotics* 10, 287–295. doi:10.1007/s11370-017-0232-y
Truong, X. T., and Ngo, T. D. (2019). "An Integrative Approach of Social Dynamic Long Short-Term Memory and Deep Reinforcement Learning for Socially Aware Robot Navigation," in Long-Term Human Motion Prediction Workshop ICRA 2019, Montreal, QC, May 20-24, 2019.
Truong, X. T., Ou, Y. S., and Ngo, T.-D. (2016). "Towards Culturally Aware Robot Navigation," in 2016 Ieee International Conference on Real-Time Computing and Robotics (Ieee Rcar) (Angkor Wat, Cambodia: IEEE), 63–69. doi:10.1109/rcar.2016.7784002
Tsai, C.-E., and Oh, J. (2020). "A Generative Approach for Socially Compliant Navigation," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 2160–2166. doi:10.1109/ICRA40945.2020.9197497
Tsoi, N., Hussein, M., Fugikawa, O., Zhao, J. D., and Vázquez, M. (2020). SEAN-EP: A Platform for Collecting Human Feedback for Social Robot Navigation at Scale. ArXiv201212336 Cs.
Turner, J. C. (1981). Towards a Cognitive Redefinition of the Social Group. *Cah. Psychol. Cogn. Psychol. Cogn.* 1, 93–118.
Van Den Berg, J., Guy, S. J., Lin, M., and Manocha, D. (2011). "Reciprocal N-Body Collision Avoidance," in *Robotics Research*. Berlin: Springer, 3–19. doi:10.1007/978-3-642-19457-3_1
van Toll, W., Grzeskowiak, F., Gandía, A. L., Amirian, J., Berton, F., Bruneau, J., et al. (2020). "Generalized Microscropic Crowd Simulation Using Costs in Velocity Space," in Symposium on Interactive 3D Graphics and Games (ACM), 1–9. doi:10.1145/3384382.3384532
Vasconez, J. P., Guevara, L., and Cheein, F. A. (2019). "Social Robot Navigation Based on HRI Non-verbal Communication," in Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 957–960. doi:10.1145/3297280.3297569
Vasquez, D. (2016). "Novel Planning-Based Algorithms for Human Motion Prediction," in 2016 IEEE International Conference on Robotics and Automation (ICRA) (Stockholm, Sweden: IEEE), 3317–3322. doi:10.1109/ICRA.2016.7487505
Vega, A., Cintas, R., Manso, L. J., Bustos, P., and Núñez, P. (2020). "Socially-Accepted Path Planning for Robot Navigation Based on Social Interaction Spaces," in Fourth Iberian Robotics Conference: Advances in Robotics, Robot 2019, 644–655. doi:10.1007/978-3-030-36150-1_53
Vega, A., Manso, L. J., Cintas, R., and Núñez, P. (2019a). Planning Human-Robot Interaction for Social Navigation in Crowded Environments. *Adv. Phys. Agents* 855, 195–208. doi:10.1007/978-3-319-99885-5_14
Vega, A., Manso, L. J., Macharet, D. G., Bustos, P., and Núñez, P. (2019b). Socially Aware Robot Navigation System in Human-Populated and Interactive Environments Based on an Adaptive Spatial Density Function and Space Affordances. *Pattern Recognition Lett.* 118, 72–84. doi:10.1016/j.patrec.2018.07.015
Vega-Magro, A., Manso, L. J., Bustos, P., and Nunez, P. (2018). "A Flexible and Adaptive Spatial Density Model for Context-Aware Social Mapping: Towards a More Realistic Social Navigation," in 2018 15th International Conference on Control, Automation, Robotics and Vision, ICARCV, 1727–1732. doi:10.1109/ICARCV.2018.8581304
Vemula, A., Muelling, K., and Oh, J. (2018). "Social Attention: Modeling Attention in Human Crowds," in Proceedings - IEEE International Conference on Robotics and Automation (IEEE), 4601–4607. doi:10.1109/icra.2018.8460504
Walters, M. L., Dautenhahn, K., Te Boekhorst, R., Kheng Lee Koay, K. L., Kaouri, C., Woods, S., et al. (20052005). "The Influence of Subjects' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction experiment," in ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 347–352. doi:10.1109/ROMAN.2005.1513803
Weina Ge, W., Collins, R. T., and Ruback, R. B. (2012). Vision-based Analysis of Small Groups in Pedestrian Crowds. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 1003–1016. doi:10.1109/tpami.2011.176
Xia, F., Zamir, A. R., He, Z., Sax, A., Malik, J., and Savarese, S. (2018). "Gibson Env: Real-World Perception for Embodied Agents," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Salt Lake City: IEEE), 9068–9079. doi:10.1109/CVPR.2018.00945
Yan, Z., Duckett, T., and Bellotto, N. (2017). "Online Learning for Human Classification in 3D LiDAR-Based Tracking," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Basel, Switzerland: IEEE), 864–871. doi:10.1109/IROS.2017.8202247
Yang, C.-T., Zhang, T., Chen, L.-P., and Fu, L.-C. (2019). "Socially-aware Navigation of Omnidirectional mobile Robot with Extended Social Force Model in Multi-Human Environment," in Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, Bari, Italy, 6-9 Oct. 2019 (Bari, Italy: IEEE), 1963–1968. doi:10.1109/smc.2019.8913844
Yang, F., and Peters, C. (2019). "Social-aware Navigation in Crowds with Static and Dynamic Groups," in 2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), Vienna, Austria, 4-6 Sept. 2019 (Vienna, Austria: IEEE), 1–4. doi:10.1109/vs-games.2019.8864512
Yang, F., Yin, W., Bjorkman, M., and Peters, C. (2020). "Impact of Trajectory Generation Methods on Viewer Perception of Robot Approaching Group Behaviors," in 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 509–516. doi:10.1109/RO-MAN47096.2020.9223584
Yao, C., Liu, C., Liu, M., and Chen, Q. (2021). Navigation in Multi-Agent System with Side Preference Path Optimizer. *IEEE Access* 9, 113944–113953. doi:10.1109/ACCESS.2021.3104470
Yao, X., Zhang, J., and Oh, J. (2019). Following Social Groups: Socially Compliant Autonomous Navigation in Dense Crowds. ArXiv Prepr. ArXiv191112063, 2–5.
Yokoyama, N., Ha, S., and Batra, D. (2021). Success Weighted by Completion Time: A Dynamics-Aware Evaluation Criteria for Embodied Navigation. ArXiv210308022 Cs.
Yoon, H.-J., Widdowson, C., Marinho, T., Wang, R. F., and Hovakimyan, N. (2019). Socially Aware Path Planning for a Flying Robot in Close Proximity of Humans. *ACM Trans. Cyber-phys. Syst.* 3, 1–24. doi:10.1145/3341570
Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., et al. (2011). Evaluating Human-Robot Interaction. *Int. J. Soc. Robotics* 3, 53–67. doi:10.1007/s12369-010-0081-8
Zhang, Y., Zhang, C.-H., and Shao, X. (2021). User Preference-Aware Navigation for mobile Robot in Domestic via Defined Virtual Area. *J. Netw. Comput. Appl.* 173, 102885. doi:10.1016/j.jnca.2020.102885
Zhong, S. W. J., Leong, L. W., and Ang, M. H., Jr. (2019). "Socially-acceptable Walking Parameters for Wheelchair Automation," in Proceedings of the Ieee 2019 9th International Conference on Cybernetics and Intelligent Systems (Cis) Robotics, Automation and Mechatronics (Ram) (Cis & Ram 2019), 193–197. doi:10.1109/cis-ram47153.2019.9095791
Zhou, Y., Wu, H., Cheng, H., Qi, K., Hu, K., Kang, C., et al. (2021). Social Graph Convolutional LSTM for Pedestrian Trajectory Prediction. *IET Intell. Transp. Syst.* 15, 396–405. doi:10.1049/itr2.12033
Zou, X., Sun, B., Zhao, D., Zhu, Z., Zhao, J., and He, Y. (2020). Multi-modal Pedestrian Trajectory Prediction for Edge Agents Based on Spatial-Temporal Graph. *IEEE Access* 8, 83321–83332. doi:10.1109/ACCESS.2020.2991435

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/frobt.2021.703807

Pose Generation for Social Robots in Conversational Group Formations

Marynel Vázquez, Alexander Lew, Eden Gorevoy and Joe Connolly*

Department of Computer Science, Yale University, New Haven, CT, United States

We study two approaches for predicting an appropriate pose for a robot to take part in group formations typical of social human conversations subject to the physical layout of the surrounding environment. One method is model-based and explicitly encodes key geometric aspects of conversational formations. The other method is data-driven. It implicitly models key properties of spatial arrangements using graph neural networks and an adversarial training regimen. We evaluate the proposed approaches through quantitative metrics designed for this problem domain and via a human experiment. Our results suggest that the proposed methods are effective at reasoning about the environment layout and conversational group formations. They can also be used repeatedly to simulate conversational spatial arrangements despite being designed to output a single pose at a time. However, the methods showed different strengths. For example, the geometric approach was more successful at avoiding poses generated in nonfree areas of the environment, but the data-driven method was better at capturing the variability of conversational spatial formations. We discuss ways to address open challenges for the pose generation problem and other interesting avenues for future work.

Keywords: human–robot interaction (HRI), group conversations, F-Formations, spatial behavior analysis, proxemics

OPEN ACCESS

Edited by:
Bilge Mutlu, University of Wisconsin-Madison, United States

Reviewed by:
Mary Ellen Foster, University of Glasgow, United Kingdom
Florian Georg Jentsch, University of Central Florida, United States

****Correspondence:***
Marynel Vázquez
marynel.vazquez@yale.edu

Specialty section:
This article was submitted to Human–Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *30 April 2021*
Accepted: *08 October 2021*
Published: *17 January 2022*

Citation:
Vázquez M, Lew A, Gorevoy E and Connolly J (2022) Pose Generation for Social Robots in Conversational Group Formations. Front. Robot. AI 8:703807. doi: 10.3389/frobt.2021.703807

1 INTRODUCTION

In this work, we study how to generate appropriate poses for social robots to take part in conversational group formations with users. This problem is important because people naturally establish these spatial formations with social robots when conversing with them (Hüttenrauch et al., 2006; Kuzuoka et al., 2010; Vázquez et al., 2015a; Karreman et al., 2015; Bohus et al., 2017). Further, people expect robots to conform to these formations when adapting to changes to group members (Vázquez et al., 2017; Yang et al., 2017).

Although it is common to model conversational spatial behavior with discriminative models of group formations (Truong and Ngo, 2017; Vázquez et al., 2017; Hedayati et al., 2019; Barua et al., 2020; Swofford et al., 2020), we approach the problem of predicting a pose for a robot in a group conversation with generative models. These models can directly output poses for the robot based on the social context of the interaction and spatial constraints imposed by the environment, for example, due to small objects such as tables or bigger structures such as walls. An illustrative example is provided in **Figure 1**.

In this work, we explore two approaches for generating spatial behavior: a model-based, geometric approach that explicitly encodes important properties of conversational group formations as often discussed in the social psychology literature (Kendon, 1990), and a data-driven adversarial approach that, once trained, implicitly encodes these properties. While our geometric approach builds directly in some cases on prior work, to the best of our knowledge, no prior effort has explored generating suitable spatial behavior for conversations subject to spatial constraints due to the environment

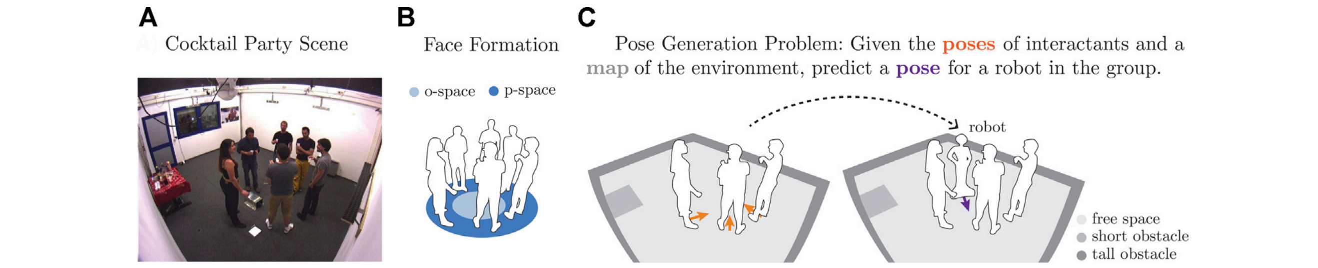


FIGURE 1 | Conversational spatial arrangement from the Cocktail Party dataset **(A)**, spatial formation typical of group conversations **(B)**, and problem setup **(C)**.

layout. By studying these two methods, this work contributes not only novel approaches, but also better understanding of how model-based and data-driven solutions for spatial reasoning in Human–Robot Interaction (HRI) complement each other.

We evaluated the proposed approaches quantitatively and qualitatively in relation to expected spatial behavior. Also, we conducted an online evaluation to gather human opinions about each method's performance when applied to situated human-robot interactions. Our results show that incorporating spatial constraints into models for pose generation is beneficial. Further, we show that the proposed methods can be used to effectively model a nonparametric distribution of poses for conversational groups. In practice, we find that considering this distribution when generating an appropriate pose can lead to better results than predicting a single pose directly. Interestingly, the human evaluation suggested that the geometric approach was more effective than the data-driven method when applied to small groups such as dyadic interactions, but the data-driven approach was better for groups with four to six interactants. We discuss ways to address this disparity. Lastly, we demonstrate the applicability of the proposed approaches for simulating conversational spatial arrangements.

2 BACKGROUND

Before explaining how the proposed methods work, the next sections provide a brief introduction to conversational group formations from a social psychology perspective and introduce Graph Neural Networks (GNNs) from a message-passing point of view. The former description is important for contextualizing the proposed geometric approach for pose generation and for understanding the rationale behind several of the metrics used in our evaluation. The latter primer on GNNs aids in understanding the proposed data-driven approach.

2.1 Conversational Group Formations

During human conversations, people often position and orient themselves in special spatial patterns known as Face Formations (F-Formations) (Kendon, 1990). F-Formations are characterized by people being nearby one another such that they can communicate easily. Also, interactants tend to direct their lower bodies toward one another or toward a common focus of attention for the conversation. These behaviors lead to spatial arrangements where individuals typically have equal, direct, and exclusive access to a common space. The formations keep groups as separate units from other close interactions.

Figure 1A depicts an example F-Formation from the Cocktail Party dataset (Zen et al., 2010), a computer vision dataset that is often used for evaluating group detection approaches based on human spatial behavior (Ricci et al., 2015; Setti et al., 2015). As illustrated in **Figure 1B**, the interior region of an F-Formation is known as its *o-space*. The area where people stand around the o-space is the *p-space*. Later in this article, we refer to these terms when formally describing geometric properties of F-Formations.

2.2 Graph Neural Networks

In this work, we use the message-passing framework for GNNs proposed by Gilmer et al. (2017), Battaglia et al. (2018) to design our data-driven pose generator method. In contrast to more traditional algorithms for reasoning about graphs, GNNs allow for learning representations, the structure of entities, and relations from graph data. Consider a graph $\mathcal{G} = (\mathbf{u}, V, E)$, where the vector $\mathbf{u}$ is a global attribute (or feature) for the graph, the set $V = \{\mathbf{v}_i\}_{i=1:n}$ corresponds to features for the graph's vertices, and $E = \{(\mathbf{e}_k, r_k, s_k)\}_{k=1:m}$ is the set of edge features $\mathbf{e}_k$ with (r_k, s_k) being the indices of the nodes connected to the edge. Then, a Graph Network block (GN block)—the basic element of a GNN—can be used to transform a graph $\mathcal{G}$ into an updated graph $\mathcal{G}' = (\mathbf{u}', V', E')$ via three steps. First, the edge features are updated. Second, the node features are updated, potentially using aggregated edge information. Third, the global attribute for the graph is updated, perhaps using node and edge information as well. Because these operations are implemented via differentiable functions, as further detailed below, the GN block can be integrated as a module into more complex neural network models.

In this work, we are concerned with using GNNs to compute vector representations for fully connected social interaction graphs that describe conversations. These graphs have a global attribute $\mathbf{u}$, corresponding to contextual information for the interaction, such as the layout of the physical environment. The graphs' node features encode pose information for the interactants, but they have no relevant edge features. Thus, applying the GN block computation to them consists of two main steps: updating the nodes features and then updating the global graph attribute. The updated global graph attribute is used to represent the graph in downstream tasks. Mathematically, we can express the two key GN block operations as follows:

$$\mathbf{v}'_i = \phi^v(\mathbf{v}_i) \tag{1}$$

$$\bar{\mathbf{v}}' = \rho^{v\rightarrow u}(\{\mathbf{v}'_i\}_{i=1:\,n}) \tag{2}$$

$$\mathbf{u}' = \phi^u(\bar{\mathbf{e}}', \bar{\mathbf{v}}', \mathbf{u}) \tag{3}$$

where the *update* functions $\phi^v(\cdot)$, $\phi^u(\cdot)$ and the *aggregate* function $\rho^{v\rightarrow u}(\cdot)$ are differentiable functions. In general, the aggregate function should take a variable number of arguments so that the GN block is suitable for processing different graphs. In addition, $\rho^{v\rightarrow u}(\cdot)$ is often implemented as a symmetric mathematical function (such as element-wise summations or maximum) because it is common for graph nodes to lack a natural order. Note that **Equations 1–3** are similar to deep set operations (Qi et al., 2017; Zaheer et al., 2017). Indeed, deep sets are sometimes regarded as a specialization of GNNs (Battaglia et al., 2018).

3 RELATED WORK

Experimental HRI work has validated the idea that spatial formations typical of human–human conversations naturally emerge in human-robot interactions (Hüttenrauch et al., 2006; Kuzuoka et al., 2010; Karreman et al., 2015; Vázquez et al., 2015a, 2017). In turn, this research led to work on recognizing F-Formations in robotics, such as methods geared toward improving robot navigation (Rios-Martinez et al., 2011), generating multimodal nonverbal robot behavior (Vázquez et al., 2017), helping recognize the beginning and ending of human-robot interactions (Gaschler et al., 2012), joining groups (Barua et al., 2020), and other approaches for service robots (Hedayati et al., 2019; Swofford et al., 2020). Oftentimes, prior work on F-Formation detection in robotics builds on mathematical models of human F-Formations from the computer vision community, for example, (Cristani et al., 2011; Setti et al., 2013; Setti et al., 2015; Vascon et al., 2014). In a similar manner, mathematical models from computer vision inspired the proposed geometric approach for generating poses for a robot in a conversation and motivated a variety of evaluation metrics in this work.

Several methods for generating spatial behavior representative of F-Formations have been proposed in HRI. For example, Vázquez et al. (2016) explored reinforcement learning for adapting the pose of a robot during conversations. Morales et al. (2014) proposed a method for a robot to walk side-by-side to a human. In addition, other work has investigated methods for robots to approach F-Formations (Shi et al., 2011; Truong and Ngo, 2017; Yang and Peters, 2019; Yang et al., 2020a). Among these methods, that of Yang and Peters (2019) is closest to our work because they explore generative adversarial networks to predict appropriate robot navigation behavior. Similar to this prior work, we are interested in modeling spatial behavior during group conversations; different to it, though, we make predictions without temporal information and subject to environmental spatial constraints, for example, nearby walls and objects.

Close to our work, Swofford et al. (2020) used a neural network model to detect F-Formations. We build on this effort because we use a similar network architecture to handle variable group sizes. Interestingly, we make an explicit connection between this prior work—which was inspired by deep sets—and GNNs following (Gilmer et al., 2017; Battaglia et al., 2018). It is worth nothing that the idea of representing interactions with graphs (as described in **Section 2.2**) is inspired by foundational work on detecting F-Formations (Hung and Kröse, 2011; Vascon et al., 2014) and a long history of applications of graph theory to social network analysis (Scott, 1988; Borgatti et al., 2009; Hamilton et al., 2017).

Among prior work that has used GNNs to reason about situated social interactions, that of Yang et al. (2020b) is perhaps the closest prior effort. While their work aimed to classify human behavior in group social encounters, we instead use GNNs to model properties of spatial formations and predict an interactant's pose within an adversarial neural network framework. Battaglia et al. (2018) and Hamilton (2020) discuss broader applications of GNNs, which are beyond the scope of this paper.

Another important related work is that of Yang et al. (2017), which proposed an approach for a robot to position itself relative to humans during a group conversation. Because this approach builds on geometric properties of F-Formations, we consider it as a baseline for the proposed methods in our evaluation.

There has also been interest in generating appropriate spatial behavior for social agents within the virtual agent community. For example, Jan and Traum (2007) considered the problem of computing agents' positions in order to create circular group formations. We also consider circular groups in this work, although these arrangements are often idealistic, as shown in our experiments. In addition, Pedica and Högni Vilhjá lmsson (2010) proposed an approach to generate human-like motion for virtual characters based on the territorial organization of social situations, including F-Formation systems. Their approach used a combination of low-level reactive behaviors to control the pose of social avatars as they move in virtual worlds. Similar to this work, we consider social norms as a driving factor when generating spatial behavior for robots and when evaluating the results of the proposed methods. Different to this prior effort, we do not expect a user to provide pose commands for the social agent of interest; rather, we study the problem of automatically generating suitable poses for a robot in a conversation.

4 GENERATING APPROPRIATE POSES DURING CONVERSATIONS

We contribute two approaches to generate an appropriate pose for a social robot in a group conversation. The key novelty of these methods stems from considering environmental spatial constraints along with the pose of other interactants upon making a prediction. These methods are both generative models, capable of representing the distributions of suitable poses via a discrete set of samples.

4.1 Problem Statement

Consider a social robot in a human environment in which there are other people with whom the robot wants to establish a situated conversation. We formulate the problem of generating an appropriate pose for the robot to sustain the conversation as follows: let $C = \{ < \mathbf{x}_i, \theta_i > \mid 1 \le i \le P\}$ be the social *context* of the interaction encoded by the poses of the P people with whom the robot wants to converse, where $\mathbf{x}_i = [x_i \;\; y_i]^T$ is their position on the ground, and θ_i is their body orientation. In addition, let M be a metric two-dimensional (2D) *map* with semantic labels for the physical environment surrounding the group. The labels encode the probability of occupancy, such as occupied space by a "small or movable object" or a "tall barrier" like a wall. Then, the goal is to compute a pose $\bar{p} = <\mathbf{x}, \theta>$ for the robot to take part in the conversational group given C and M, as illustrated in **Figure 1C**. The generated pose should preserve the spatial structure of the group, that is, their F-Formation. Also, the pose should be such that the robot does not collide with objects according to the map, as well as does not violate social norms such as personal space.

4.2 A Geometric Approach for Pose Generation

One way to compute a viable pose for a robot to take part in a conversational group is to explicitly formalize key geometric properties of its expected spatial behavior. To this end, we first consider the fact that F-Formations often have a circular shape because of people's tendency to position in a way such that they can see and monitor one another during conversations (Kendon, 1990). The circular shape not only defines an expected distribution for people's locations but also guides their body orientations toward the center of their group's o-space. Second, we consider the fact that the agent should not be in an occupied location and should not violate other people's personal space.

Based on the above properties, we propose a three-step algorithm for computing a pose $\bar{p} = <\mathbf{x}, \theta>$ given the context C and map M:

1) Fit circular shape to the context poses. We represent the geometric shape of the group formation parametrically with a 2D circle or ellipse fitted to the context C (as illustrated in **Figures 1A,B**). The edge of the shape represents the p-space of the F-Formation, whereas its interior corresponds to the o-space. Intuitively, fitting an ellipse should be preferable to fitting a circle because of the variability of human spatial behavior. However, we sometimes default to using circles because fitting ellipses requires at least 5 points.

   To fit a circle, we consider three cases. First, if the context has a single individual, $|C| = 1$, then we assume that the center of the circle is d units in front of the individual, in the direction of its transactional segment. This means that the o-space of the group is defined by the circle with a center at $\mathbf{c} = \mathbf{x}_1 + d\,[\cos(\theta_1) \;\; \sin(\theta_1)]^T$ and a radius of d. The distance d has been defined in the literature as the *stride* parameter of mathematical F-Formation models (Cristani et al., 2011). Second, if the context has two individuals, $|C| = 2$, then we assume that the center of their group's o-space is in between them because face-to-face spatial arrangements are common for dyads. This means that the center of the circle is given by $\mathbf{c} = (\mathbf{x}_1 + \mathbf{x}_2)/2$, and its radius is $\|\mathbf{x}_1 - \mathbf{x}_2\|/2$. Third, if the context has at least three people, $|C| >= 3$, then we fit a circle to their locations using orthogonal distance regression (Boggs and Rogers, 1990), which tends to be more robust to potential errors in the location measurements than ordinary least squares.

   To fit an ellipse to the location of the interactants in C, we follow the direct fitting approach by Halíř and Flusser (1998). We found this approach to be fast in comparison to iterative approaches and more robust than that of Fitzgibbon et al. (1996) when $|C| = 5$.
2) Compute the robot's location. We view the problem of computing a suitable location for the robot given the fitted circular shape, the context C, and map M as an optimization problem. The key factor in this formulation is the loss function, which we define as a weighted sum of three components that penalize for deviations from the fitted circular shape (ℓ_c), close proximity to other individuals (ℓ_p), and positioning in nonfree areas of the environment (ℓ_f). Formally:

$$\ell(\mathbf{x}) = \lambda_c \ell_c(\mathbf{x}) + \lambda_p \ell_p(\mathbf{x}) + \lambda_f \ell_f(\mathbf{x}) \quad (4)$$

   where $\lambda_c, \lambda_p, \lambda_f \in \mathbb{R}^+$ control the effect of each penalty. The first component ℓ_c corresponds to the perpendicular distance from $\mathbf{x}$ to the fitted circle or ellipse. The second component ℓ_p penalizes violations to personal space: $\ell_p(\mathbf{x}) = \sum_{i=1}^{P} \mathcal{N}(\mathbf{x}; \mathbf{x}_i, I\sigma)$, where $\mathcal{N}$ denotes a normal distribution with mean $\mathbf{x}_i$ and variance σ. Lastly, ℓ_f in **Eq. 4** is a penalty for the input location corresponding to a nonfree cell of the map M. **Figures 2A–E** illustrate these different components for the loss, where the map has been smoothed to avoid positions too close to nonfree cells.

   While one could use brute-force search to find a minima of **Eq. 4** around the context C, we propose to minimize the loss using Powell's conjugate direction method (Powell, 1964), a popular optimization algorithm. This method does not require derivatives, which is convenient for this optimization because computing the orthogonal distance to an ellipse, as needed by the ℓ_c penalty, is a nontrivial problem for which we use an iterative method. See Uteshev and Yashina (2015) and Uteshev and Goncharova (2018) for a discussion on the point-to-ellipse problem.
3) Compute the robot's orientation. We finally set θ such that the robot orients toward the center of the fitted circular shape, corresponding to the expected center of the o-space.

4.3 A Data-Driven Adversarial Approach for Pose Generation

Another way to approach the problem of generating a suitable pose for a robot in a conversation is to leverage generative data-driven methods. In particular, we explore using Wasserstein Generative Adversarial Networks (WGAN), originally proposed by Arjovsky et al. (2017), to produce poses that conform to measured

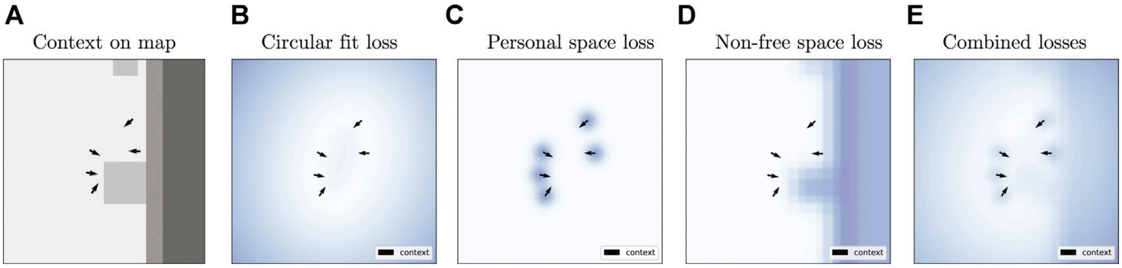


FIGURE 2 | Example losses for the Geometric approach on a sample from the Cocktail Party dataset. From left to right: **(A)** social context on an environment map with free space (light gray color), short obstacles (medium gray), and tall obstacles (darker gray); **(B)** circular fit loss; **(C)** personal space loss; **(D)** penalty associated with nonfree cells of the environment map; and **(E)** weighted sum of those three losses. The arrows indicate the pose of the interactants. Brighter values in the loss plots correspond to lower cost.

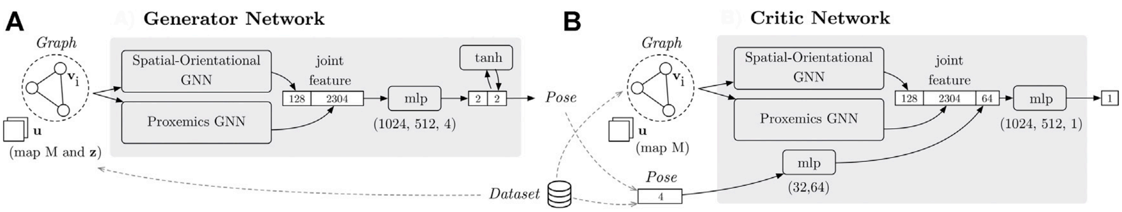


FIGURE 3 | The generator **(A)** and the critic **(B)** process the information in the social interaction graph via two GNNs. One GNN reasons about the spatial–orientational arrangement of the group (encoded in the vertex features $\mathbf{v}_i$). The other GNN reasons about proxemics based on the interactant's positions (encoded in the vertices) and the map (encoded in the global attribute **u**). Note that the global attribute for the graph input to the generator also includes the latent variable **z**. The "mlp" blocks are multilayer perceptrons.

characteristics of F-Formations. This type of data-driven model is composed of two neural networks: a *generator G*, which we use to predict the desired pose $\bar{p}$; and a *discriminator D*, which helps discern generated poses from poses in the true data. Note that for WGANs, D is often called the *critic* because the network is not trained to classify, but outputs a real value; here, we use the terms interchangeably to help readers familiar with adversarial networks follow our explanation.

Without loss of generality, let us represent the pose of a social agent $< \mathbf{x}, \theta >$ as a 4D row vector $\mathbf{p} = [x\ y\ \cos(\theta)\ \sin(\theta)]$ so that we do not have to worry about θ wrapping around the $(-\pi, \pi]$ interval. Also, assume that we have a dataset $\mathcal{D} = \{< C^j, M^j, \mathbf{p}^j >\}$ with ideal poses $\mathbf{p}$ for a social robot given a corresponding context C and map M. Our goal with the WGAN is to then train the generator and discriminator networks using $\mathcal{D}$. Formally, the WGAN objective can be expressed as a minimax game:

$$\min_G \max_D \mathbb{E}_{\mathbf{p}\sim\mathbb{P}_r}[D(\mathbf{p}|C, M)] - \mathbb{E}_{\bar{\mathbf{p}}\sim\mathbb{P}_g}[D(\bar{\mathbf{p}}|C, M)] \qquad (5)$$

where we have conditioned the discriminator D on the corresponding context and map data for the sampled pose, following the formulation for Conditional Generative Adversarial Networks by Mirza and Osindero (2014). The discriminator (or critic) in **Eq. 5** should be in the set of 1-Lipschitz functions, which we implement via a gradient penalty added to the loss in **Eq. 5** per (Gulrajani et al., 2017). Lastly, $\mathbb{P}_r$ in **Eq. 5** is the real data distribution induced by $\mathcal{D}$, and $\mathbb{P}_g$ is the distribution implicitly defined by the generator G: $\bar{\mathbf{p}} = G(\mathbf{z}|C, M)$, with the latent variable $\mathbf{z} \sim p(\mathbf{z})$ coming from a simple prior (e.g., a standard normal distribution in this work).

We propose a novel two-stream architecture for the generator and discriminator networks (**Figure 3**). This architecture is driven by our knowledge of the problem domain—we take advantage of inductive biases (in terms of relational and spatial structure) to facilitate learning. The next sections provide more details.

4.3.1 The Generator Network

Figure 3A describes how the generator predicts a pose $\bar{\mathbf{p}}$ given a social interaction graph $\mathcal{G}$ as input. The nodes of the graph correspond to pose features $\mathbf{v}_i = [x_i\ y_i\ \ \cos(\theta_i)\ \sin(\theta_i)]$. The graph's global attribute $\mathbf{u}$ is a tensor with dimensions $3 \times h \times w$. The first two channels correspond to the map $M \in \mathbb{R}^{2\times h\times w}$, which represents occupancy by tall and short barriers in its first and second channels, respectively. The last channel of $\mathbf{u}$ corresponds to the latent variable $\mathbf{z}$.

One processing stream of the generator reasons about the graph $\mathcal{G}$ focusing on the *spatial–orientational arrangement* of the interactants (i.e., the information in the node features) using a GNN that operates in the same spirit as deep sets (Qi et al., 2017; Zaheer et al., 2017)—similar to the "context transform" proposed by Swofford et al. (2020). Another parallel stream processes the graph focusing on *proxemics*, that is, how interactants use space in relation to the

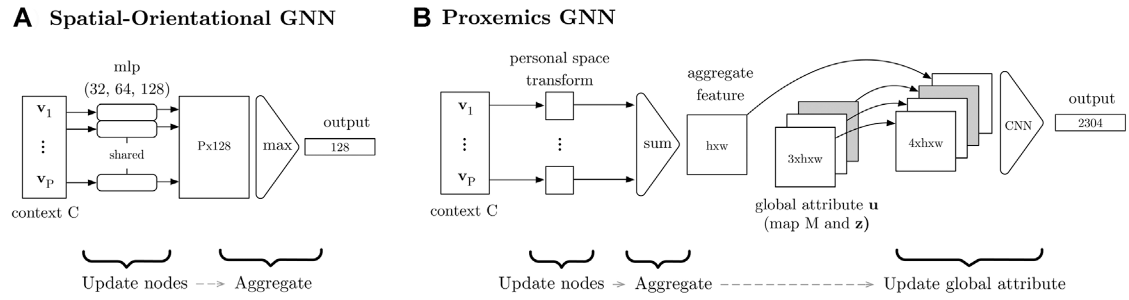


FIGURE 4 | Graph neural networks used in the generator. The "mlp" component in **(A)** is a multi-layer perceptron and "CNN" in **(B)** is a convolutional neural network.

environment (Hall, 1966). This stream is a GNN that uses 2D convolutional layers to reason about two types of spatial relationships: the shape of the formation based on the location of interactants (encoded in the vertices of $\mathcal{G}$) and the location of nearby objects relative to the group (based on the map in the graph's global attribute).

The generator concatenates the vector representations that result from the two computation streams and then transforms the data through a two-layer perceptron (with ReLU transformations) and one additional linear layer. This results in a 4D output vector, whose first two elements correspond to the position of the output pose $\bar{\mathbf{p}}$. The last two elements are the cosine and sine of the robot's orientation, which are constrained to lie in $(-1, 1)$ through a final hyperbolic tangent transformation applied to these elements.

The next sections explain how the parallel streams of the generator network are implemented. More implementation details are provided in the **Supplementary Material**.

4.3.1.1 Spatial–orientational GNN

Figure 4A illustrates the architecture of the spatial–orientational component of the generator. The network is a GN block that aggregates position and orientation information from the group: $\mathbf{u}_1' = \rho_1^{v\to u}(\{\phi_1^v(\mathbf{v}_i)\}_{i=1:P})$, where the update function ϕ_1^v corresponds to a multilayer perceptron (with ReLU activations) applied to the vertex features $\mathbf{v}_i$, and the aggregate function $\rho_1^{v\to u}$ is max pooling. Comparing these operations with **Eqs. 1–3**, this GN block can be thought of as having a trivial ϕ^u function in **Eq. 3** that simply returns the aggregate feature for the nodes.

4.3.1.2 Proxemics GNN

Figure 4B depicts the generator's proxemics component, which is also a GN block. First, the GN block updates the node features by creating a 2D tensor $\mathbf{v}_i' = \phi_2^v(\mathbf{v}_i) \in \mathbb{R}^{h\times w}$ that represents the personal space of the interactant i using a simple Gaussian blob. That is, $\mathbf{v}_i'$ is a matrix of the same width and height as the map M, where each cell corresponds to a physical location in the world and has a value equal to the probability density of a normal distribution centered at the location of the interactant $[x_i\, y_i]$. Second, the GN block aggregates the updated node features $\bar{\mathbf{v}}' = \sum_{i=1:P}\mathbf{v}_i'$ using element-wise summation. Third, the global attribute is updated by concatenating $\mathbf{u}$ (with the map and latent variable $\mathbf{z}$) with the aggregated personal space representation $\bar{\mathbf{v}}'$, resulting in a tensor in $\mathbb{R}^{4\times h\times w}$. The latter tensor is then processed by a three-layered convolutional neural network with ReLU activation, and the result is finally flattened into a vector representation $\mathbf{u}_2'$ for this stream. Note that the node update and aggregate functions used by this GNN lead to a representation similar to the personal space loss used for the Geometric approach (and illustrated in **Figure 2C**). However, the network is not told explicitly how to reason about this data; instead, it needs to figure this out through the adversarial training regimen implemented with the critic.

4.3.2 The Critic Network

We implement the critic in a similar fashion to the generator, with two data processing streams. The main difference is that instead of getting an input graph whose global attribute contains a latent variable $\mathbf{z}$, the global graph attribute $\mathbf{u} = M$ in this case. Also, the critic gets an additional input pose $\mathbf{p}$, which may come from the dataset $\mathcal{D}$ or from the output of the generator. This pose is processed in a third parallel stream, as illustrated in **Figure 3B**, using a two-layer perceptron with ReLU activations. The three-vector-representations output by the two GNNs and the pose streams are concatenated and finally projected into a scalar value. The **Supplementary Material** provides more details on the GNNs and this last transformation.

4.4 Generating a Distribution of Poses

Both the geometric and WGAN approach described previously can be used to generate a nonparametric distribution of poses for conversational group formations. This is useful in two ways: (1) it can help identify multiple poses that may be suitable for a given conversational group, and (2) it can help overcome predictions that are not optimal, perhaps because of local minima. The latter is particularly important for the Geometric approach because its output is subject to the initial location provided to its optimization routine. Also, computing a distribution can be useful for the WGAN because its generator is not guaranteed to output an ideal pose given an arbitrary input latent vector $\mathbf{z}$. Indeed, the neural network is trained to model the distribution of the real data, not a single pose.

Generating a distribution of poses with both approaches is trivial. For the Geometric approach, we can start its optimization step from different initial locations around the context C, predicting various locations for the agent. Then, we can compute suitable orientations for each of the locations as explained in **Section 4.2**. For the WGAN, we simply need to run the generator multiple times using different latent variables as input. Once a nonparametric distribution of poses is computed, we can choose a single pose as output if desired. For example, in our evaluation in **Sections 5–6**, we do this by searching for a mode of the predicted locations using the mean shift algorithm (Comaniciu and Meer, 2002) and then simply outputting the pose in the distribution that is closest to this mode. We also tried more involved approaches such as computing a mode for the angle of the pose as well using a von Mises kernel density estimator (Fisher, 1995). However, the former approach gave similar or better performance in practice and reduced the number of hyperparameters that we needed to consider in our implementation, facilitating future reproducibility.

5 EVALUATION ON THE COCKTAIL PARTY DATASET

This section first evaluates the proposed approaches quantitatively with respect to different metrics that describe key properties of F-Formations and desired output poses. Then, we discuss the results qualitatively.

5.1 Datasets

We used the Cocktail Party dataset (Zen et al., 2010) to evaluate the proposed approaches. The dataset consists of approximately 30 min of interaction data. It includes 320 frames with conversational group annotations and pose information for six individuals who took part in a Cocktail Party event, as shown in **Figure 1A**. While the original dataset provides head orientation for each of the individuals based on automatic tracking methods, our evaluation used manually annotated body orientations (Vázquez et al., 2015b) as θ for the pose of interactants. Reasoning about body orientation instead of head orientation preserves consistency with the theory of F-Formations (Kendon, 1990). In addition to this data, we manually created an environment map for the Cocktail Party scene with labels for "free space," space occupied by "tall objects" (through which social interactions are unlikely), and space occupied by "short objects" (like the table in the room). Areas outside of the Cocktail Party room were labeled as having "unknown" occupancy in the map and were treated as occupied space in practice.

We split the group annotations from the Cocktail Party dataset into two sets: training (80%) and testing (20%). The test set included 31 frames with group annotations at the beginning of the Cocktail Party sequence, 31 frames in the middle, and 31 more at the end; the training set was composed of the other frames with group annotations.[1] The latter groups were then used to create a dataset $\mathcal{D}^{CP}_{train} = \{<C, M, \mathbf{p}>\}$ of 1,394 examples with corresponding contexts C, map M, and example ground truth pose $\mathbf{p}$ for a robot. The map for these examples had 24 × 24 cells and a resolution of 0.25 m per cell. They were a cropped section (generated with subpixel accuracy) of the full environment layout, covering an area of approximately 3-m radius around the context C. The ground ruth pose in the examples corresponded to the position and orientation of one member of the group who was excluded from the context. Using the test groups, we created a similar dataset $\mathcal{D}^{CP}_{test}$ for evaluating the proposed models, where $|\mathcal{D}^{CP}_{test}| = 347$.

We also created a dataset of simulated F-Formations using 15 environment layouts from the iGibson simulation environment (Shen et al., 2020). For each environment, we first created a 2D layout intersecting the 3D geometry of the world with planes parallel to the ground, as illustrated in **Figures 5A, B**. Using the layout, we then manually created an environment map with the same labels and resolution of the Cocktail Party environment map and automatically generated circular groups with two to six people in free areas of the environment following a simple rule-based procedure. This resulted in 34,405 simulated examples, each with a corresponding environment map, context and example ground truth pose for the robot. **Figure 5C** shows one sample from this dataset.

Upon preliminary testing of the data-driven method, we realized that the WGAN significantly benefited from many diverse examples. Thus, we further augmented the dataset of simulated groups by warping the data using a small amount of horizontal and vertical stretch as well as random rotations. This resulted in an expanded dataset of 60,365 simulated examples in total, which we used to train some variations of the data-driven model in this evaluation. The **Supplementary Material** provides more details about the data generation process used to create simulated F-Formations in iGibson environments.

5.2 Pose Generation Methods

The present evaluation considered variations of the proposed methods and a recent baseline for robot pose generation in F-Formations, which does not use information about the surrounding physical environment. To the best of our knowledge, no prior work has considered variability in the environment of F-Formations when generating suitable poses for an interactant. All methods were implemented in Python.

Baseline method: We implemented the pose generation method by Yang et al. (2017). As with our model-based approach, this method seeks a circular spatial pattern but without explicitly accounting for environment characteristics. Instead, the existing social context alone determines the generated pose, which is computed as follows: First, any pair of individuals in the context is used to define a mutual circular region. Second, all pairwise centers are averaged to compute the center coordinate of the o-space, to which the new member faces. The minimum and maximum distances of individuals to the common center demarcate the p-space of the group, the annular zone that interacting peers occupy (as in **Figure 1B**). Finally, bisecting the largest gap between adjacent neighbors identifies the new member's position within the group.

Geometric methods: We evaluated the Geometric approach proposed in **Section 4.2** considering two cases. In one case, the method generates a single pose using an initial location for its optimization step that is within a 3 × 3 m region around the center of

[1]Splitting the dataset in this manner minimized overlap between the training and testing data given the temporal correlation of the data.

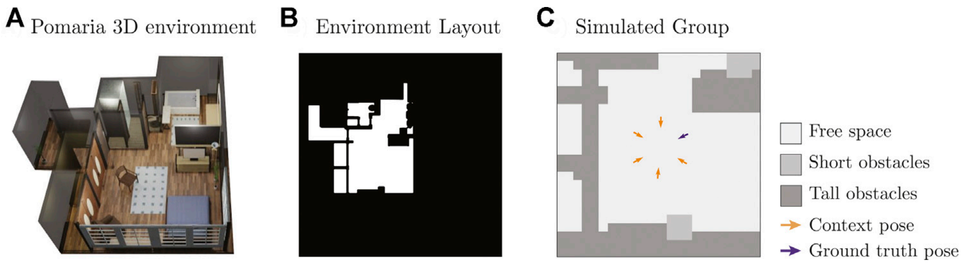


FIGURE 5 | (A) Example 3D environment from iGibson (Shen et al., 2020), **(B)** environment layout generated from the 3D environment, and **(C)** simulated sample on a cropped section of the layout.

the given context C. In the other case, we run the method multiple times to model a distribution of poses and then use mean shift to choose an output pose (as described in **Section 4.4**). In the latter case, we initialize the method with 36 different initial values for its optimization, which are sampled uniformly in the same 3×3 m region considered in the former case. For both variations, we set the loss parameters $\sigma = 0.21$, $\lambda_p = 1.25$, $\lambda_c = 0.2$ and $\lambda_f = 0.5$ based on initial results on $\mathcal{D}^{CP}_{train}$.

Data-driven methods: We considered three variations of the proposed WGAN. First, we considered a model trained on the simulated dataset (with a small amount of angular noise applied to the orientation of the context poses to make the group arrangements more varied). Second, we evaluated a model trained on the Cocktail Party train data only (with 10% used for validation). Third, we considered a model trained like the first one and then fine-tuned on the Cocktail Party train data. In addition, we considered generating one sample pose from the generator, as well as generating a distribution of 36 poses from which we output a solution guided by mean shift (as in **Section 4.4**).

We implemented the WGAN using the PyTorch library and trained models using an NVidia GeForce RTX 2080 Ti GPU. More specifically, we used the Adam optimizer with a learning rate of 0.00002, a batch size of 32, and a weight of 10 for the WGAN gradient penalty (Gulrajani et al., 2017). During gradient descent, we weighted the training samples based on the relative distribution of group sizes in the dataset and updated the critic five times for every generator update. We trained models for at least 600 epochs and chose the best training weights through a combination of manual inspection of the generated samples and quantitative metrics on the Cocktail Party validation data.

The **Supplementary Material** provides more implementation details for the WGAN. Also, it describes results for several other variations of the WGAN that we explored in this work, but that resulted in no major improvement. For example, we considered a model that only had information about free space, instead of multiple map labels.

5.3 Quantitative Metrics

We considered a range of metrics that describe F-Formations and social norms in regard to spatial behavior:

- Deviations from fitted circle or ellipse (Circ. Fit). We measure the perpendicular distance from a generated pose to a circle or ellipse that has been fitted to the context C. The circle or ellipse is fitted following the same considerations described in **Section 4.2** for the proposed Geometric approach.
- Individual is not on free space (Not Free). We compute how often the location of a generated pose falls within a nonfree cell in the environment map M. The values for this metric ranged in [0, 1] because of subpixel cropping of the maps.
- Violations to personal space (Per. Space). We compute the number of cases in which the distance between the generated pose $\bar{p} = \langle \mathbf{x}, \theta \rangle$ and the pose of another member of the group $\langle \mathbf{x}_j, \theta_j \rangle$ is less than a personal space threshold, $\|\mathbf{x} - \mathbf{x}_j\| < \delta$. We use a threshold of $\delta = 0.68$ m based on real-world data of interpersonal distances in Italy (Sorokowska et al., 2017), because the Cocktail Party data were originally captured in that country.
- Violations to intimate space (Int. Space). Similar to personal space, we compute the number of cases in which the distance between the generated pose and another group member j is less than an intimate space threshold, $\|\mathbf{x} - \mathbf{p}_j\| < \rho$. We use $\rho = 0.42$ m based on Sorokowska et al. (2017).
- Distance to group's o-space center (Center Dist.). Let $\mathbf{x}_i$ and θ_i be the location and body orientation of a social agent (human or robot) in a conversational group. Prior work, such as those of Cristani et al. (2011) and Setti et al. (2013, 2015), has proposed to compute the o-space center of an F-Formation as follows:

$$\bar{\mathbf{o}} = \frac{1}{P}\sum_{i=1}^{P}\mathbf{o}_i = \frac{1}{P}\sum_{i=1}^{P}\left(\mathbf{x}_i + d\begin{bmatrix}\cos(\theta_i)\\ \sin(\theta_i)\end{bmatrix}\right) \tag{6}$$

  where P is the number of interactants in the group, and $\mathbf{o}_i$ is a proposed o-space center for member i. Thus, we measure alignment with an ideal F-Formation model as the average distance between the group's o-space center and the o-space center proposals for individual members: $CenterDist = \frac{1}{P}\sum_{i=1}^{P}\|\bar{\mathbf{o}} - \mathbf{o}_i\|$. For the parameter d, needed to compute $\bar{\mathbf{o}}$ in **Eq. 6**, we use $d = 0.72$ as it minimizes $\sum_{g=1}^{K}\sum_{i=1}^{P^g}\|\bar{\mathbf{o}}_g - \mathbf{o}_i\|^2$, considering all K ground-truth groups for the Cocktail Party dataset (see Vázquez (2017) for the derivation).
- Individual occludes another interactant (Occ. Other). Ideally, the generated poses should not be in front of other interactants, as this would prevent them from having direct access to the o-space and exclude them from the group. To identify these situations, we check if

TABLE 1 | Results on the Cocktail Party test set.

	Method	Circ. Fit	Not Free	Per. Space	Int. Space	Center Dist.	Occ. Other	Is Occ.
1	Ground Truth	0.35 ± 0.24	0.01 ± 0.09	0.48 ± 0.56	0.00 ± 0.00	0.27 ± 0.10	0.00 ± 0.00	0.00 ± 0.00
2	Yang	3.02 ± 21.51	0.26 ± 0.43	0.28 ± 0.62	0.00 ± 0.05	1.20 ± 9.44	0.00 ± 0.00	0.02 ± 0.14
3	Geometric	0.33 ± 0.29	0.00 ± 0.05	0.01 ± 0.13	0.00 ± 0.00	0.30 ± 0.24	0.01 ± 0.27	0.09 ± 0.28
4	WGAN (iG)	0.33 ± 0.29	0.07 ± 0.23	0.38 ± 0.61	0.11 ± 0.33	0.45 ± 0.15	0.03 ± 0.17	0.01 ± 0.12
5	WGAN (CP)	0.28 ± 0.23	0.02 ± 0.13	0.73 ± 0.71	0.31 ± 0.49	0.46 ± 0.12	0.10 ± 0.40	0.06 ± 0.23
6	WGAN (iG, CP)	0.31 ± 0.23	0.03 ± 0.16	0.68 ± 0.64	0.22 ± 0.41	0.45 ± 0.11	0.12 ± 0.32	0.01 ± 0.12
7	Geometric*	0.29 ± 0.28	0.00 ± 0.05	0.01 ± 0.13	0.00 ± 0.00	0.30 ± 0.24	0.00 ± 0.05	0.07 ± 0.26
8	WGAN* (iG)	0.33 ± 0.28	0.07 ± 0.23	0.36 ± 0.59	0.10 ± 0.29	0.45 ± 0.15	0.03 ± 0.18	0.01 ± 0.09
9	WGAN* (CP)	0.29 ± 0.23	0.02 ± 0.13	0.72 ± 0.68	0.32 ± 0.49	0.46 ± 0.12	0.12 ± 0.40	0.04 ± 0.20
10	WGAN* (iG, CP)	0.31 ± 0.22	0.03 ± 0.15	0.66 ± 0.65	0.21 ± 0.41	0.45 ± 0.11	0.11 ± 0.32	0.02 ± 0.13

*Each row shows μ ± σ for each of the metrics described in **Section 5.3** (lower is better). Models without * output a single pose, whereas those with * output a distribution of 36 poses from which we chose a single pose (guided by the mode of the distribution) as final output. "(iG)" models were trained on simulated data using iGibson environment maps, "(CP)" indicates training with Cocktail Party train data, and "(iG,CP)" corresponds to pretraining with simulated data and then fine-tuning on Cocktail Party train data. The best results (for which there are no significant differences) are highlighted in gray per column—see the text for statistical analyses.*

the generated pose is in between another interactant in the group and the o-space center $\bar{\mathbf{o}}$. The center $\bar{\mathbf{o}}$ is computed as in **Eq. 6** while excluding the generated pose, because a bad prediction could skew significantly $\bar{\mathbf{o}}$.

– Individual is behind another interactant (Is Occ.). This metric is similar to the prior metric, but we invert the roles of the generated pose and an interactant's pose for which we compute the occlusion.

The first three metrics correspond to each of the losses considered by the Geometric approach and thus serve to validate that the method was working as expected. In addition, these metrics are useful to evaluate whether the data-driven method behaved in a similar manner. The occlusion metrics are inspired by the visibility constraints from Vázquez et al. (2015b) and Setti et al. (2015), and the personal and intimate space metrics signal potential violations to social norms. Lastly, the Center Dist. metric serves to evaluate the combined effect of position and orientation prediction. The metrics are inspired by ideal models of F-Formations, but real-world data may not perfectly satisfy all assumptions set forth by the metrics. Thus, we report values for ground truth test data in our results as a reference for comparison.

5.4 Quantitative Results

Table 1 presents the results on the Cocktail Party test set. As a reference, the first row shows the values for the metrics using ground truth poses from the test set (which were removed to create the context C input to the pose generation methods shown in **Table 1**).

Unless noted otherwise, we analyzed the results for the quantitative metrics using restricted maximum likelihood (REML) analyses considering method (10 levels, each one corresponding to a row of **Table 1**) as main effect and Example ID from the Cocktail Party test set as random effect. The results for the Circ. Fit metric indicated a significant effect of method ($F[9, 3114] = 5.50$, $p < 0.0001$). A Tukey honestly significant difference (HSD) *post hoc* test showed that the baseline method by Yang et al. (2017) led to significantly higher Circ. Fit values than the other methods. The baseline performed poorly because its o-space representation is the average of all circles fitted to pairs of group members. Thus, a single pair can heavily bias the position of the generated interactant. For example, we often observed this bias when the difference between the orientations of a pair of individuals in the context was small, which resulted in a circle with a disproportionately long radius. There were no other significant pairwise differences for the Circ. Fit results, suggesting that the proposed methods were able to effectively capture the circularity of F-Formations.

An REML analysis on the Not Free metric indicated that there were significant differences by Method ($F[9, 3114] = 64.72$, $p < 0.0001$). The *post hoc* test showed that the baseline method by Yang et al. (2017) resulted in significantly more poses generated in occupied cells of the environment map than all other methods. This was expected because the baseline did not consider the environment map in its calculations. The only other pairwise differences for the Not Free metric were the results for rows 4 and 8 in **Table 1**, which were low but significantly higher than the results for rows 1, 3, 5, 7, and 9. As a reference, rows 4 and 8 corresponding to the WGAN trained on iGibson-simulated data led to 24/347 and 22/347 examples for which the Not Free metric was greater than 0.5. Meanwhile, the ground truth values had 3/347 instances in this category, and the Geometric approach led to only one such case in the Cocktail Party test set.

We also found significant differences for violations to personal space ($p < 0.0001$) and intimate space ($p < 0.0001$) using REML analyses, as well as using Poisson generalized mixed linear models with a log link function. In terms of Per. Space, a Tukey HSD *post hoc* test showed that the Geometric approach (rows 3 and 7 in **Table 1**) led to significantly lower number of personal space violations than all other methods, followed by the Yang baseline (row 2) and the WGAN trained on simulated data using iGibson environments (rows 4 and 8). Also, the WGAN trained or fine-tuned on Cocktail Party train data led to significantly higher violations to Per. Space than all other methods. In terms of Int. Space, best results were obtained with the Ground Truth poses (row 1), the Yang baseline (row 2), and the Geometric approaches (rows 3 and 7). These methods had significantly fewer intimate space violations than all other methods. Further, the WGAN trained on simulated data (rows 4 and 8) was significantly better in terms of Int. Space than the other WGAN variations (rows 5, 6, 9, and 10). The WGAN fine-tuned on Cocktail Party train data (rows 6 and 10) was also significantly better than the WGAN trained on these data only (rows 5 and 9).

The results for the Center Dist. metric were similar to the Circ. Fit metric: an REML analysis showed significant differences per Method ($F[9, 3114] = 2.75$, $p = 0.003$), and the *post hoc* test showed that the Yang baseline had significantly worse Center Dist. results than all the other methods.

The values for the occlusion metrics were generally low, but there were significant differences across Methods. For Occ. Other ($p < 0.0001$), the methods in rows 1–4, 7, and 8 in **Table 1** resulted in poses that led to significantly fewer occlusions than the methods in rows 5, 6, 9, and 10. For the Is Occ. metric ($p < 0.0001$), the Tukey HSD *post hoc* indicated that the Ground Truth results (row 1 in **Table 1**) were significantly lower than those for the Geometric approach (rows 3 and 7) and the WGAN trained on the Cocktail Party train data (rows 5 and 9). However, there were no significant pairwise differences between the Ground Truth results, the Yang baseline (row 2), the WGAN trained on iGibson data (rows 4 and 8), or the WGAN fine-tuned on Cocktail Party train data (rows 6 and 10).

In summary, the results in **Table 1** led to three key takeaways. First, the proposed methods worked better than the baseline in terms of the Circ. Fit, Not Free, and Center Dist. metrics. This showed the value of considering environmental spatial constraints when predicting poses for agents in conversational groups and the superiority of the proposed methods at modeling the shape of F-Formations. Second, training the WGAN on simulated data using iGibson environments turned out to be as good as or better than training on realistic Cocktail Party data only, except for the Not Free metric for which the simulated data led to slightly worse results. We attribute this result to the fact that generative adversarial models are data-hungry, and the Cocktail Party train set had only 1,394 examples (approximately 2% of the simulated dataset). Effective fine-tuning of the WGAN model on the small Cocktail Party train set proved difficult. Third, computing a distribution of poses led to slight improvements in some cases compared to predicting a single pose directly. For instance, the distribution helped slightly the WGAN model in terms of personal space violations and the Geometric approach in terms of occlusions.

5.5 Qualitative Results

We further analyzed the results from **Section 5.4** qualitatively for the baseline by Yang et al. (2017), the Geometric approach and the WGAN (trained on simulated data). **Figures 6A–E** shows example results by these methods on different group sizes. The columns are identified with the same naming convention as **Table 1**, where * in the Figure corresponds to methods that internally predicted a distribution of 36 poses.

In comparison to the baseline (Yang column), the proposed Geometric approach resulted in similar predictions when the context had one or two poses (**Figures 6A,B**). However, for bigger groups, the Geometric approach tended to model circular spatial arrangements more consistently than the baseline, resulting in poses that were better positioned or oriented with respect to the context.

In regard to the methods that computed pose distributions, **Figure 6** shows that these distributions captured different viable solutions to the pose generation problem. Interestingly, while the Geometric* approach tended to lead to more multimodal distributions than the WGAN* (iG), the data-driven method led to fewer occluded poses in these distributions. Occlusions were a problem for the Geometric approach due to local minima in its optimization step, but by predicting multiple poses, this problem was alleviated.

Figure 7 shows more difficult prediction problems, where the context poses are distributed in less circular form or are closer to physical obstacles. These cases led to poor o-space modeling for both the baseline and the Geometric approach. In particular, in **Figure 7A**, the Geometric approach fit a circular shape to the context that had a disproportionately big radius and was oriented in the wrong direction. In **Figure 7B**, the baseline by Yang et al. (2017) had trouble with pairs of poses in the context being oriented very similar to one another, which led to a generated pose that was very far away from the group. Also, in **Figure 7C**, the baseline output a generated pose in nonfree space.

In terms of the WGAN, **Figure 7B** shows that the WGAN had more trouble avoiding short obstacles than the other methods. Furthermore, **Figure 7C** shows that another failure for the WGAN was to place poses toward the center of a group. Predicting a distribution of poses in this case was useful in comparison to generating a single pose, as the distribution comprised poses in more appropriate positions relative to the context.

Despite the challenges encountered in some cases by the proposed approaches, they generally performed better than the baseline by Yang et al. (2017) both in terms of considering environmental constraints and dealing with the variability inherent in human spatial behavior. However, it was hard to evaluate the methods holistically: we did not know of a good way to combine the quantitative metrics considered in this section into a single success measure. Thus, to complement these results, we conducted a complementary, human-driven evaluation of the proposed approaches. This evaluation is presented in the following section.

6 HUMAN EVALUATION

We evaluated generated poses by the proposed geometric and data-driven approaches from a human perspective. For this evaluation, we chose a diverse set of the groups from the Cocktail Party test data. We then removed one human member of the groups, as in the prior evaluations, and computed the pose for a robot to be part of the interaction with the remaining members. The resulting spatial arrangement was rendered in a virtual scene similar to the environment of the Cocktail Party dataset. Human participants then gave us their opinion of the pose of the robot relative to the virtual humans rendered in the scenes.

This experiment followed a similar protocol to Connolly et al. (2021). The main difference is that our focus was not on evaluating the effect of different robot embodiments on human perception of conversational groups; instead, we wanted to compare the two proposed methods for pose

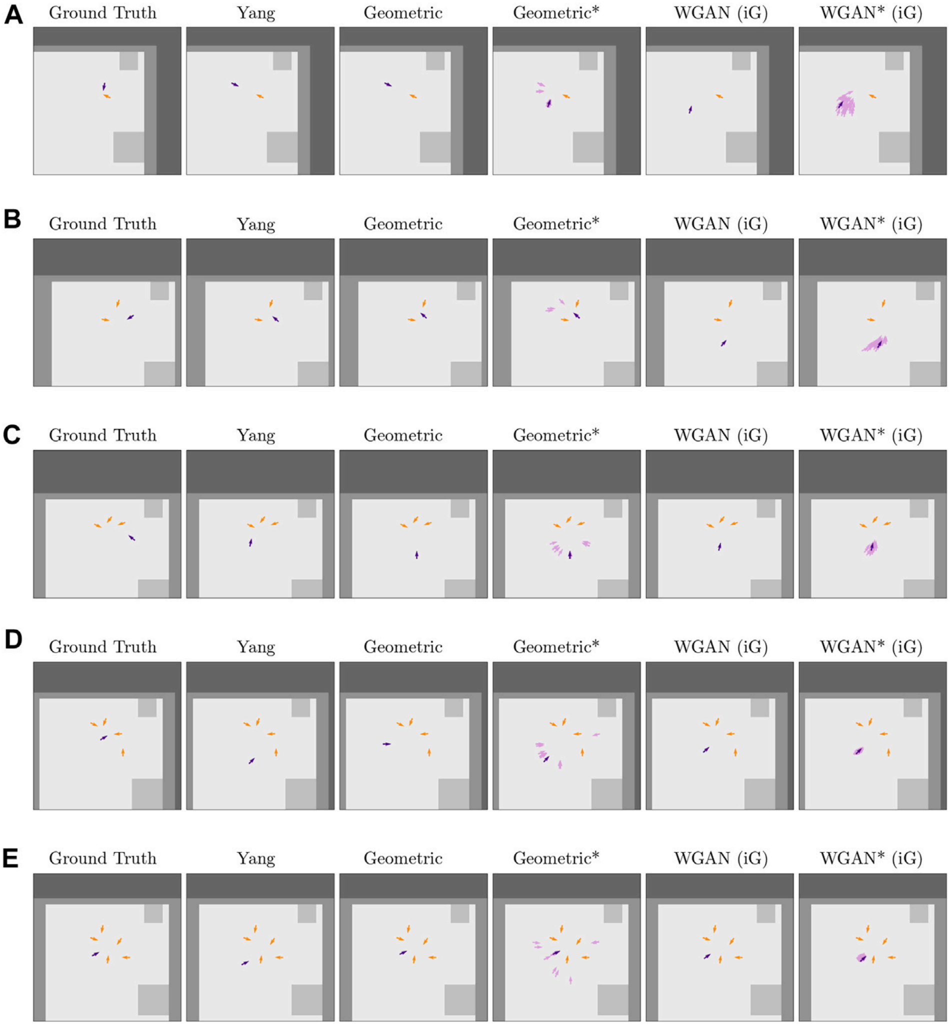


FIGURE 6 | Successful predictions for several methods (one per column) on five different problems (rows) from the Cocktail Party test set. The orange arrows correspond to the poses in the context *C*, and the purple arrows are predicted poses, except for the Ground Truth column in which the purple arrows correspond to a true pose by a group member. Note that the darker purple arrows are the final output by each method, and the lighter ones are additional predictions by the methods that computed a distribution of poses. The colors of the environment map are the same as in **Figure 5C**: "free space" is light gray, "short obstacles" is medium-intensity gray, and "tall obstacles" is darker gray. In addition, these plots show one more label for "unknown" occupancy (darkest gray color). The latter label was a result of cropping the full environment map around context poses next to the edge of the map. When computing results, "unknown" occupancy was considered as nonfree space by the Geometric approach and was aggregated with tall obstacles for the WGAN. This Figure is best viewed in color.

generation. For this reason, we focused on using a single robot embodiment for this study. The selected robot was a humanoid Pepper robot, which has an easily discernible body orientation and head. Its dimensions are similar to those of a young person.

6.1 Participants

We used Prolific to recruit a total of 60 participants (32 females and 28 males) for this human evaluation. The participants resided in the United States, were fluent or native English speakers, had normal or corrected-to-normal vision, and had an average age of 32.15 years (standard deviation [STD] = 12.57). They indicated sometimes playing video games (mean [M] = 4.32, STD = 2.14) and rarely interacting or working with a robot (M = 2.23, STD = 1.59) on 7-point responding formats (1 being the lowest rating and 7 being highest).

6.2 Experiment Design

We controlled for three main variables in this evaluation:

Method (two levels). We compared the Geometric approach (**Section 4.2**) with the WGAN approach (**Section 4.3**). Both methods computed a distribution of 36 samples from which we chose a single output pose by searching for a mode across predicted locations (as explained in **Section 4.4**). For both methods, we used the best hyperparameters found in **Section 5**. For the WGAN, in particular, we chose the model that was trained on simulated iGibson data (row 8 of **Table 1**) because, except for the Not Free metric, it led to better or similar performance than alternatives.

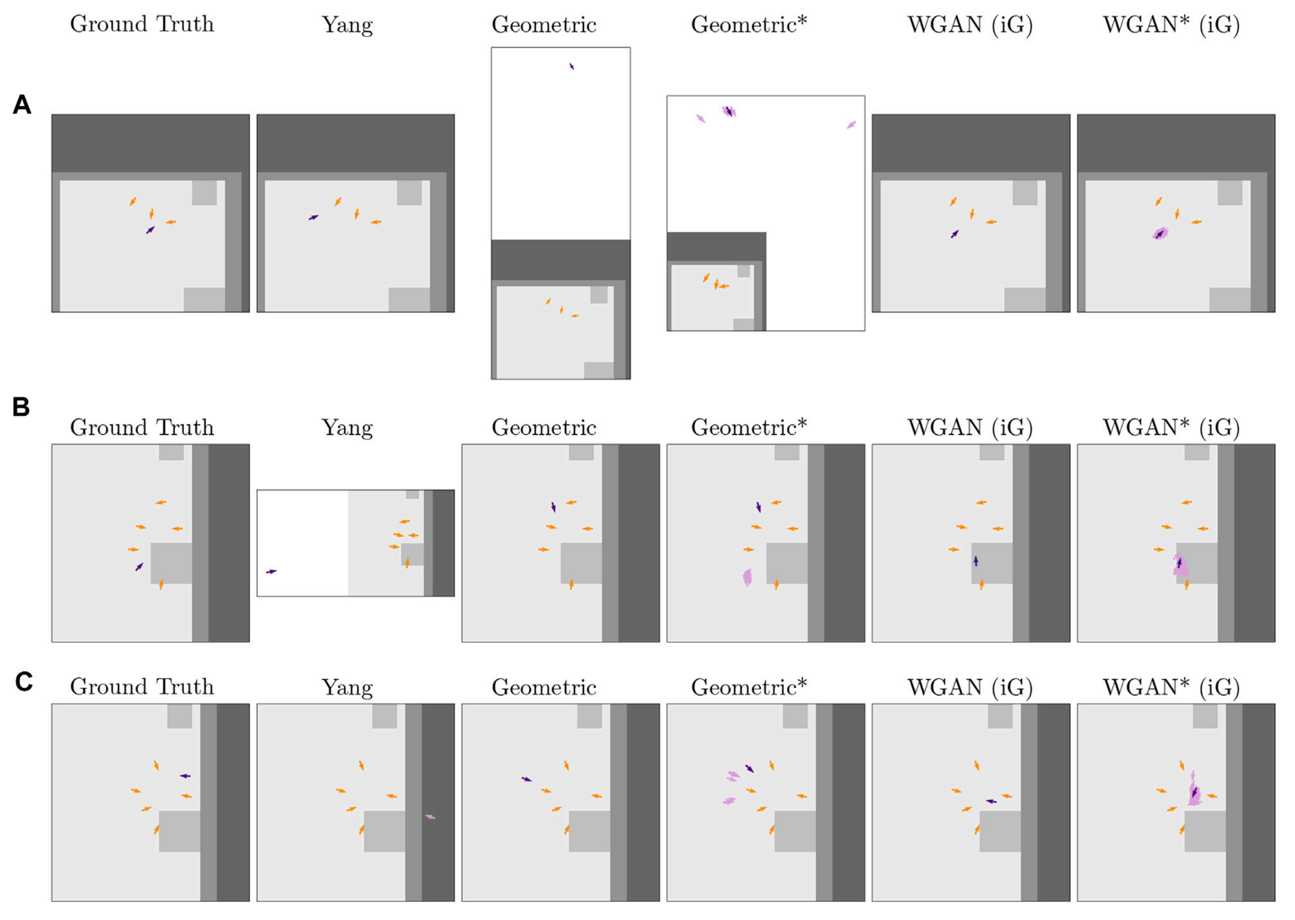


FIGURE 7 | Difficult predictions on three different problems (rows) from the Cocktail Party test set. As in **Figure 6**, the orange arrows correspond to the poses in the context *C*, the purple arrows are Ground Truth or predicted poses, and the lighter gray color in the maps corresponds to free space. White areas within a plot correspond to regions of the space out of the cropped map (considered as "unknown" occupancy by the proposed methods). This Figure is best viewed in color.

Context (10 levels). We considered 10 contexts (i.e., interactants' poses input to the methods) for each of the Group Sizes mentioned below. The contexts were chosen to try to maximize the diversity of scenarios considered in the evaluation and without looking at how well the proposed methods performed on them.

Group Size (five levels). We considered group sizes of two to six interactants (including the robot). This means that the proposed methods had as input a Context with one to five interactants.

The study was run with a mixed design using a Qualtrics online survey. The participants provided their opinion of the pose of the robot for a single Group Size (between participants) in renderings generated for all Context/Method combinations (within participants). In particular, for each combination of Context and Method, we generated two renderings that depicted the resulting interaction.[2] One rendering corresponded to a top-down view of the group, and the other was a frontal view so that the participants could easily perceive the robot's spatial positioning relative to the other interactants (as shown in **Figure 8A**).

The participants were randomly assigned to each Group Size category, resulting in all categories having at least four males or females. Renderings made for Group Sizes of three, four, and six interactants were evaluated by 12 participants each, whereas the renderings for Group Sizes of two and five interactants were evaluated by 13 and 11 participants, respectively.

6.3 Measures

The participants provided feedback about the pose of the Pepper robot on each scene shown in their survey, each of which corresponded to a given combination of Context and Method. In particular, the survey first asked them to visually identify the Pepper robot in the rendered scene. Then, it asked them to rate four statements about the robot's pose relative to the virtual humans. Example images can be seen in **Figure 8**, along with the statements that the participants had to rate using a 7-point Likert responding format from "strongly disagree" (1) to "strongly

[2]The renderings were generated as in Connolly et al. (2021), using the Unity game engine (https://unity.com/), tools from the Social Environment for Autonomous Navigation (Tsoi et al., 2020), the Microsoft Rocketbox avatar library (Gonzalez-Franco et al., 2020), and an open-source version of Pepper's Universal Robot Description File (http://wiki.ros.org/pepper_description).

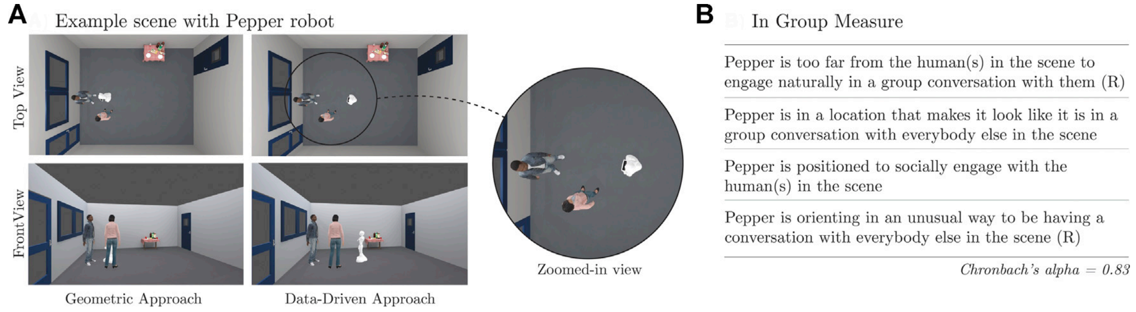


FIGURE 8 | Example scene used for the human evaluation **(A)** and statements rated by the participants about the robot's pose **(B)**. (R) indicates that the ratings were reversed before computing the "In Group" measure.

agree" (7). Following Connolly et al. (2021), we reversed the scores for negative statements and computed the correlation between them, obtaining moderate positive pairwise correlations (see **Supplementary Table S8** in the supplementary material). Cronbach α for these ratings was 0.83, above the nominal 0.7 threshold. Thus, we grouped responses into an "In Group" measure.

6.4 Procedure

Upon starting the survey, the participants completed a consent form to take part in the evaluation and provided basic demographics data (as described in **Section 6.1**). Then, the survey showed renderings of two practice scenes and asked the participants to rate the pose of the robot in them using the In Group statements from **Figure 8B**. The practice scenes depicted different Contexts than those used for the evaluation to avoid biasing participant's opinion. In one practice scene, the pose of the robot corresponded to the ground truth pose for the individual that it replaced in the Cocktail Party dataset. In the other practice scene, the robot's pose was generated by taking the ground truth pose and then reorienting the robot opposite to its group. These examples served to familiarize participants with the robot and the In Group statements used to rate its pose.

After the two practice scenes, the survey showed the real evaluation scenes. For each scene, the survey asked the participants to evaluate the pose of the Pepper robot using the In Group statements. Note that the survey for the participants who provided feedback for groups of size 4 included only 19 evaluation scenes because the Geometric approach led to positioning the robot outside of the Cocktail Party environment in one case, which we removed from our evaluation. For all other group sizes, the survey included 20 evaluation scenes as originally planned (10 contexts × 2 methods). The order of the evaluation scenes was randomized for all the participants. That is, the renderings by Method and Context were randomly interspersed with one another within a participant's survey to avoid potential ordering effects.

After rating all the scenes, the participants provided their opinion about how hard it was to complete the survey. We used these responses in pilots to improve the protocol design. The survey typically took approximately 12 min to complete, for which the participants were paid US $2.4. This protocol was approved by our local Institutional Review Board.

The **Supplementary Material** provides more details on the specific design of the online survey and shows all the renderings used in this evaluation.

6.5 RESULTS

We conducted an REML analysis on the In Group measure. In this analysis, we considered Method, Group Size, and their interaction as main effects, and Context and Participant ID as random effects. We found significant effects for Group Size (F [4, 55.19] = 3.24, p = 0.019). A Tukey HSD *post hoc* test suggested that the In Group ratings were significantly lower on groups of size 4 (M = 4.44, STD = 1.67, N = 240) than on groups of size 6 (M = 5.27, STD = 1.65, N = 240). No other significant differences were obtained by Group Size. The ratings for groups of size 2, 3, and 5 were M = 4.78 (STD = 1.37; N = 260), M = 5.14 (STD = 1.56; N = 240), and M = 4.91 (STD = 1.58; N = 220), respectively.

The REML analysis indicated that Method had a significant effect on the In Group ratings, F[1, 1115] = 10.06 (p = 0.002). A Student *t post hoc* test suggested that the data-driven approach (M = 5.01, STD = 1.44, N = 600) led to significantly higher ratings than the Geometric approach (M = 4.81, STD = 1.73, N = 600), although this difference was small (approximately 0.2 points on the 7-point scale).

There was also a significant interaction effect of Method and Group Size on the In Group measure, F[4, 1114] = 61.43 ($p <$ 0.0001). Interestingly, a Tukey HSD *post hoc* test indicated that the In Group ratings were significantly higher for the WGAN than for the Geometric approach on Group Sizes 4, 5, and 6. However, the Geometric approach led to significantly higher ratings than the data-driven method for the other Group Sizes, as illustrated in **Figure 9A**. This difference in performance by Group Sizes was observed on each of the individual components of the In Group measure, as shown in **Figure 9B**.

We looked further into the generated renderings to better understand why the methods led to different In Group values per Group Size. We noticed two trends:

1. For a Group Size of 2 and 3, the WGAN tended to place the robot farther away from the context individuals than the Geometric approach. For example, this result can be seen in **Figure 8A**. Also, additional examples can be found in **Supplementary Figures S4 and S5** in the supplementary material. For instance, for groups of size 2 in **Supplementary Figure S4**, the robot is farther away from the groups with the WGAN than with the Geometric approach

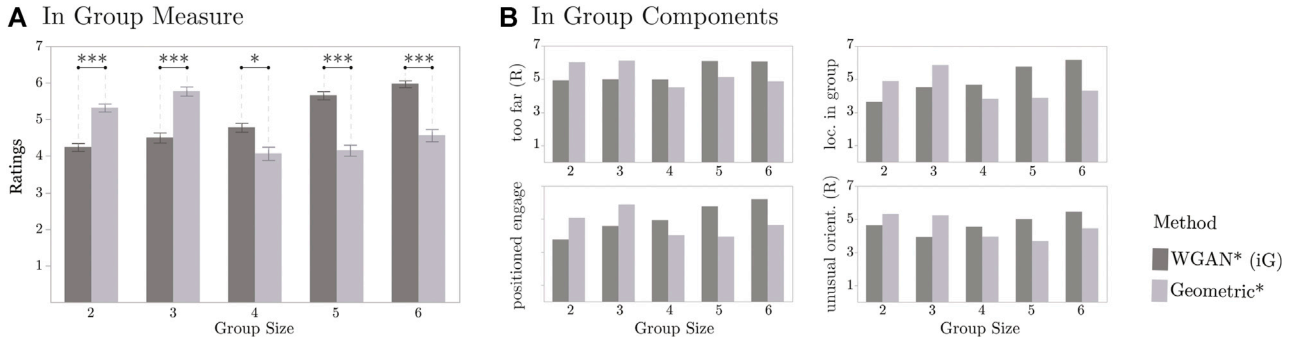


FIGURE 9 | Results from online human evaluation comparing the Geometric and data-driven approach (WGAN). In the left plot, the symbol * indicates $p < 0.05$ and *** indicates $p < 0.001$, and error bars correspond to standard. error. The right plots only show averages to illustrate the similarity among ratings.

in Contexts 2, 3, 5, 6, 7, 9, and 10. Likewise, this effect can be seen in Contexts 1, 3, 5, 6, and 8 for groups of size 3 in **Supplementary Figure S5**.

2. For Group Sizes bigger than 3, we observed in the renderings that the Geometric approach tended to place the robot more often behind individuals than the WGAN. For example, this can be seen in Contexts 4, 6, 9, and 10 for Group Size 4 in **Supplementary Figure S6** in the supplementary material. Likewise, this result can be seen in Contexts 1, 3, 4, 5, 9, and 10 for Group Size 5 in **Supplementary Figure S7**, and on Contexts 1, 3, 6, 7, 8, and 9 for Group Size 6 in **Supplementary Figure S8**. This result is a direct consequence of the hyperparameters that we chose for the model-based method. In particular, when looking at preliminary results in the Cocktail Party train dataset (as described in **Section 5**), we prioritized avoiding violations to personal and intimate space. However, this impaired the capacity of the Geometric approach to find suitable gaps for the robot in spatial arrangements that already had at least three members.

The mixed results for the In Group ratings highlight different properties of the proposed pose generation methods. First, we attribute the lower In Group ratings for the WGAN on Group Sizes 2 and 3 to the method's reliance on the training data distribution. As mentioned before, we trained the WGAN using simulated iGibson data, based on our earlier results on the Cocktail Party dataset (**Section 5**). However, these data were generated without special consideration for group size. All the groups were created by simply placing interactants along a circular arrangement in free space; we should have instead created smaller circular arrangements for smaller groups. Second, the difficulties that the Geometric approach had with Group Sizes 4, 5, and 6 speak to how challenging it is to choose suitable hyperparameters for the Geometric approach given all the many factors that matter for the pose generation problem, including proxemics, the shape of F-Formations, occlusions within groups, and the physical environment.

7 GENERATING CONVERSATIONAL GROUPS

Although we focused our work on predicting a suitable pose for a social agent in a group conversation, the proposed approaches could be reused to create entire conversational groups. These groups are constructed by invoking the proposed generative methods iteratively given a map M, the pose of an initial individual $< \mathbf{x}_0, \theta_0 >$, and the desired number of group members. After each iteration, the newest generated pose is added to the social context, which the generator subsequently takes as an input.

Figure 10 illustrates conversational groups generated by both proposed approaches using the above iterative method on two different environments: one map corresponds to the single room of the Cocktail Party dataset (**Figures 10A,B**), and the other one is drawn from the iGibson environments (**Figures 10C,D**). At each iteration of the group generation approach, the final pose output for a new interactant is selected from a distribution of 36 samples computed by the corresponding method. These samples are shown as light purple arrows in **Figure 10**. The hyperparameters for the Geometric and WGAN methods used in this section are the same parameters used for computing the results in **Sections 5 and 6**.

In general, the results for the iterative group generation task reflect prior findings. First, the Geometric approach generates poses that better respect personal space, as can be seen in the right-most column of **Figure 10**. Second, for smaller group sizes, the Geometric approach outputs poses that are more tightly positioned relative to existing group members than the WGAN; however, for bigger group sizes, the WGAN outputs poses in more circular group formations than the Geometric approach. These circular formations are prototypical of real conversational interactions, suggesting that the WGAN better identifies proxemic constraints introduced by additional interactants than the Geometric approach.

From a computational perspective, iterative invocation of the geometric method, without special care for parallelization, requires more time to output a result than the WGAN due to the inherent sequential nature of its optimization step. For example, while the WGAN might take approximately 0.08 s to make a prediction on a consumer-grade MacBook computer, the Geometric approach might take approximately 0.5 s. Owing to this higher runtime cost yet greater stability, the optimization

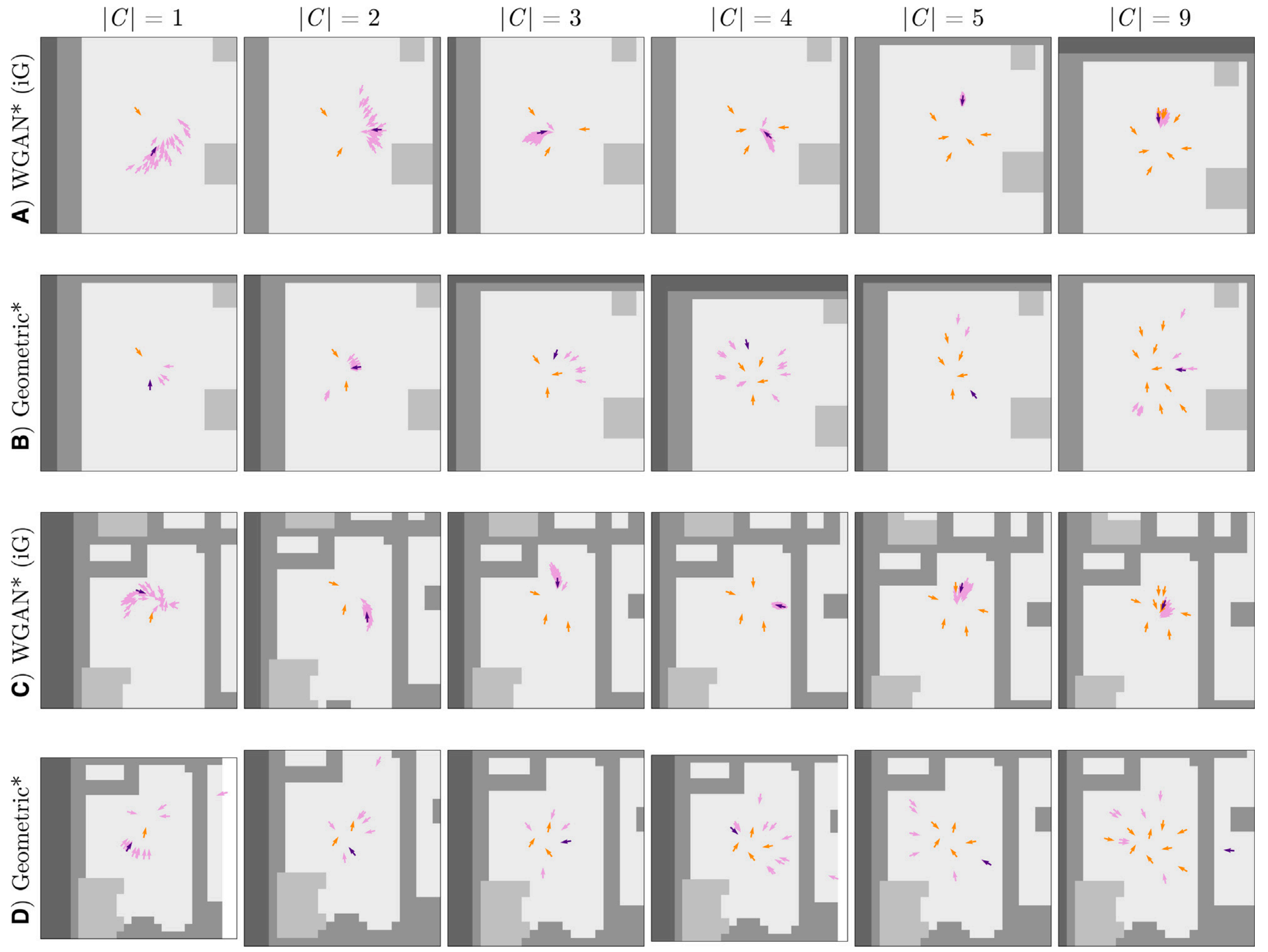


FIGURE 10 | Group generation through iterative invocations of the geometric and data-driven approaches. The plots in rows **(A)** and **(B)** show results in the map from the Cocktail Party dataset. Rows **(C)** and **(D)** demonstrate our two approaches in a map from the iGibson dataset. As in **Figure 6**, the orange arrows correspond to the poses in the context C, the purple arrows are Ground Truth or predicted poses, and the lighter gray color in the maps corresponds to free space. $|C|$ indicates the size of the context used to make a prediction per column. This Figure is best viewed in color.

approach may be used to simulate very large groups appropriate for training data-driven models in the future.

Lastly, the results also show limitations of the Geometric approach in reasoning holistically about social scenes. For example, in **Figure 10D**, the Geometric approach proposes a pose for $|C| = 9$ that is separated from the rest of the context by a wall. In contrast, the WGAN avoids placing poses far from the existing context (**Figure 10C**) without a physically based rule. This result further highlights the difficulty of handcrafting solutions to the pose generation problem, as these solutions need to effectively balance proxemics, spatial environmental constraints, and arbitrary conversational group sizes.

8 DISCUSSION

8.1 Summary of Contributions

Our work introduced two approaches for generating poses for social robots in group conversations given spatial constraints and the pose of other group members. One approach formalizes key geometric properties of spatial behavior evident in conversational groups. In this Geometric approach, generating the location of a pose is formulated as an optimization problem, whose loss function penalizes divergence from the circular shape of the existing group formation, violations of personal space, and robot placement in nonfree environmental areas. The other, data-driven approach models expected spatial behavior with a WGAN. The inputs to the generator and discriminator networks are a map of the environment and a social interaction graph, where the graph nodes correspond to the pose features of existing interactants. Our novel architectures for the generator and discriminator rely on GNNs, which reason about spatial–orientational arrangements and proxemic relationships in a more implicit manner than our Geometric approach.

We evaluated our proposed methods on the Cocktail Party dataset with metrics based on desirable properties of conversational group formations. We chose for a baseline a pose generation method that does not consider environmental characteristics.

Both of our approaches significantly outperformed the baseline method on metrics for maintaining the circular shape of the group and accounting for obstacles in the environment. This evaluation affirms the importance of considering environmental constraints in addition to interactants' poses when generating spatial behavior for social agents.

The quantitative evaluation also informed model selection for a second evaluation, which compared our two pose generation approaches from a human perspective. Study participants assessed poses generated by our two proposed approaches in virtual scenes. With respect to an In Group measure, the Geometric approach generated superior poses for groups of three or four interactants, whereas the data-driven approach scored better for larger groups. The contrasting strengths of our two approaches further reinforce the complexity of pose generation in social applications: an optimal solution must respect spatial constraints from both the environment and other group members while also considering human expectations for behavior in a variety of scenarios.

In addition to the above contributions, this work explored using the proposed pose generation methods to simulate conversational groups of different sizes. We are excited about the potential of this application to enhance robotics simulations for HRI, like SEAN (Tsoi et al., 2020), as the proposed methods could be used as a practical mechanism to add human–robot social interactions to virtual environments. This could allow the community to further study social robot navigation (Mavrogiannis et al., 2021) or advance our understanding of proxemics and human perception of spatial patterns of behavior in HRI (Li et al., 2019; Connolly et al., 2021).

8.2 Limitations and Future Work

Our work is limited in several ways, which we consider avenues for future work. First, we did not find a clear winner between the proposed pose generation methods. The Geometric approach led to best quantitative metrics, but according to human ratings, it did not perform as well as the data-driven method with bigger groups. While we believe that in the long term the data-driven approach is more likely to succeed than the Geometric approach because it has more flexibility to reason about the intricacies of human spatial behavior, it is heavily dependent on the availability of significant amounts of realistic data. Thus, future work could explore creating better datasets for pose generation subject to environmental spatial constraints and reevaluate the WGAN on such datasets. One interesting idea in this respect is leveraging the Geometric approach to augment the training data used for the data-driven method.

Second, we focused on predicting a suitable pose for a robot given the location and orientation of interactants, but one could consider additional input features for the context in the future, such as motion data. Adding this information to the Geometric approach may require additional special considerations, but providing more input features to the data-driven method is easier. For example, we could adjust the architecture of the spatial–orientational GNN used in the generator and critic to take on more input features per interactant and thus allow the networks to reason about this additional information.

Third, our work is limited in that our evaluation of the proposed approaches considered simulated interactions only. We have not yet evaluated the methods on real-world human robot interactions. In the future, we would like to study the effectiveness of the proposed methods to enable robots to adapt their pose during situated group conversations, as interactants move or come and go. We would also like to explore using the proposed methods for enabling robots to join nearby group conversations subject to physical environmental constraints.

Fourth, we often assumed in this work that robots should behave in similar ways to humans. However, prior work suggests that robot embodiment may affect the way in which people interpret robot spatial behavior in HRI (Connolly et al., 2021). Thus, future work should investigate whether the proposed methods are suitable for different types of robots, especially those that are less anthropomorphic than Pepper.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://gitlab.com/interactive-machines/spatial_behavior/genff.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Yale University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All the authors contributed to the implementation of the proposed methods and their evaluation. They also contributed to writing this article.

FUNDING

This work was supported by the National Science Foundation (NSF), Grant No. (IIS-1924802). The findings and conclusions in this article are those of the authors and do not necessarily reflect the views of the NSF. EG was also supported by the Yale Hahn Scholars program.

ACKNOWLEDGMENTS

Thanks to Jeacy Espinoza for helping create the environment map for the Cocktail Party dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2021.703807/full#supplementary-material

REFERENCES

Arjovsky, M., Chintala, S., and Bottou, L. (2017). "Wasserstein Generative Adversarial Networks," in Proceedings of the 34th International Conference on Machine Learning, 214–223.

Barua, H. B., Pramanick, P., Sarkar, C., and Mg, T. H. (2020). Let Me Join You! Real-Time F-Formation Recognition by a Socially Aware Robot. arXiv preprint arXiv:2008.10078. doi:10.1109/ro-man47096.2020.9223469

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., et al. (2018). Relational Inductive Biases, Deep Learning, and Graph Networks. arXiv preprint arXiv:1806.01261.

Boggs, P. T., and Rogers, J. E. (1990). Orthogonal Distance Regression. *Contemp. Mathematics* 112, 183–194. doi:10.1090/conm/112/1087109

Bohus, D., Andrist, S., and Horvitz, E. (2017). "A Study in Scene Shaping: Adjusting F-Formations in the Wild," in Proceedings of the 2017 AAAI Fall Symposium: Natural Communication for Human-Robot Collaboration.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network Analysis in the Social Sciences. *Science* 323, 892–895. doi:10.1126/science.1165821

Comaniciu, D., and Meer, P. (2002). Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Trans. Pattern Anal. Machine Intell.* 24, 603–619. doi:10.1109/34.1000236

Connolly, J., Tsoi, N., and Vázquez, M. (2021). "Perceptions of Conversational Group Membership Based on Robots' Spatial Positioning: Effects of Embodiment," in Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 372–376.

Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Del Bue, A., et al. (2011). Social Interaction Discovery by Statistical Analysis of F-Formations. *BMVC* 2, 4. doi:10.5244/c.25.23

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge University Press.

Fitzgibbon, A. W., Pilu, M., and Fisher, R. B. (1996). "Direct Least Squares Fitting of Ellipses," in Proceedings of 13th International Conference on Pattern Recognition (IEEE), 253–257. doi:10.1109/icpr.1996.546029

Gaschler, A., Jentzsch, S., Giuliani, M., Huth, K., de Ruiter, J., and Knoll, A. (2012). "Social Behavior Recognition Using Body Posture and Head Pose for Human-Robot Interaction," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE), 2128–2133. doi:10.1109/iros.2012.6385460

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). "Neural Message Passing for Quantum Chemistry," in International Conference on Machine Learning (PMLR), 1263–1272.

Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., et al. (2020). The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Front. Virtual Reality* 1, 1–23. doi:10.3389/frvir.2020.561558

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). "Improved Training of Wasserstein gans," in Advances in neural information processing systems, 5767–5777.

Halíř, R., and Flusser, J. (1998). "Numerically Stable Direct Least Squares Fitting of Ellipses," in International Conference in Central Europe on Computer Graphics and Visualization. WSCG, 125–132.

Hall, E. T. (1966). *The Hidden Dimension*, 609. Garden City, NY: Doubleday.

Hamilton, W. L. (2020). Graph Representation Learning. *Synth. Lectures Artif. Intelligence Machine Learn.* 14, 1–159. doi:10.2200/s01045ed1v01y202009aim046

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. arXiv preprint arXiv:1706.02216.

Hedayati, H., Szafir, D., and Andrist, S. (2019). "Recognizing F-Formations in the Open World," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 558–559. doi:10.1109/hri.2019.8673233

Hung, H., and Kröse, B. (2011). "Detecting F-Formations as Dominant Sets," in Proceedings of the 13th international conference on multimodal interfaces, 231–238. doi:10.1145/2070481.2070525

Hüttenrauch, H., Eklundh, K. S., Green, A., and Topp, E. A. (2006). "Investigating Spatial Relationships in Human-Robot Interaction," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE), 5052–5059.

Jan, D., and Traum, D. R. (2007). "Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversations," in Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, 1–3. doi:10.1145/1329125.1329142

Karreman, D., Ludden, G., van Dijk, B., and Evers, V. (2015). "How Can a Tour Guide Robot's Orientation Influence Visitors' Orientation and Formations?" in Proceeding of 4th International Symposium on New Frontiers in Human-Robot Interaction.

Kendon, A. (1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters*, 7. Cambridge: CUP Archive.

Kuzuoka, H., Suzuki, Y., Yamashita, J., and Yamazaki, K. (2010). "Reconfiguring Spatial Formation Arrangement by Robot Body Orientation," in 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 285–292. doi:10.1109/hri.2010.5453182

Li, R., van Almkerk, M., van Waveren, S., Carter, E., and Leite, I. (2019). "Comparing Human-Robot Proxemics Between Virtual Reality and the Real World," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 431–439. doi:10.1109/hri.2019.8673116

Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Steinfeld, A., Trautman, P., et al. (2021). Core Challenges of Social Robot Navigation: A Survey. arXiv preprint arXiv:2103.05668.

Mirza, M., and Osindero, S. (2014). Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784.

Morales, Y., Kanda, T., and Hagita, N. (2014). Walking Together: Side-By-Side Walking Model for an Interacting Robot. *J. Human-Robot Interaction* 3, 50–73. doi:10.5898/jhri.3.2.morales

Pedica, C., and Vilhjálmsson, H. (2010). Spontaneous Avatar Behavior for Human Territoriality. *Appl. Artif. Intelligence* 24, 575–593. doi:10.1080/08839514.2010.492165

Powell, M. J. D. (1964). An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives. *Comput. J.* 7, 155–162. doi:10.1093/comjnl/7.2.155

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). "Pointnet: Deep Learning on point Sets for 3d Classification and Segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 652–660.

Ricci, E., Varadarajan, J., Subramanian, R., Rota Bulo, S., Ahuja, N., and Lanz, O. (2015). "Uncovering Interactions and Interactors: Joint Estimation of Head, Body Orientation and F-Formations from Surveillance Videos," in Proceedings of the IEEE International Conference on Computer Vision, 4660–4668. doi:10.1109/iccv.2015.529

Rios-Martinez, J., Spalanzani, A., and Laugier, C. (2011). "Understanding Human Interaction for Probabilistic Autonomous Navigation Using Risk-Rrt Approach," in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE), 2014–2019. doi:10.1109/iros.2011.6094496

Scott, J. (1988). Social Network Analysis. *Sociology* 22, 109–127. doi:10.1177/0038038588022001007

Setti, F., Lanz, O., Ferrario, R., Murino, V., and Cristani, M. (2013). "Multi-Scale F-Formation Discovery for Group Detection," in 2013 IEEE International Conference on Image Processing (IEEE), 3547–3551. doi:10.1109/icip.2013.6738732

Setti, F., Russell, C., Bassetti, C., and Cristani, M. (2015). F-formation Detection: Individuating Free-Standing Conversational Groups in Images. *PloS One* 10, e0123783. doi:10.1371/journal.pone.0123783

Shen, B., Xia, F., Li, C., Martín-Martín, R., Fan, L., Wang, G., et al. (2020). iGibson, a Simulation Environment for Interactive Tasks in Large Realistic Scenes. arXiv preprint arXiv:2012.02924.

Shi, C., Shimada, M., Kanda, T., Ishiguro, H., and Hagita, N. (2011). Spatial Formation Model for Initiating Conversation. *Proc. Robotics: Sci. Syst. VII.*, 305–313. doi:10.15607/rss.2011.vii.039

Sorokowska, A., Sorokowski, P., Hilpert, P., Cantarero, K., Frackowiak, T., Ahmadi, K., et al. (2017). Preferred Interpersonal Distances: a Global Comparison. *J. Cross-Cultural Psychol.* 48, 577–592. doi:10.1177/0022022117698039

Swofford, M., Peruzzi, J., Tsoi, N., Thompson, S., Martín-Martín, R., Savarese, S., et al. (2020). Improving Social Awareness Through Dante: Deep Affinity Network for Clustering Conversational Interactants. *Proc. ACM Hum.-Comput. Interact.* 4, 1–23. doi:10.1145/3392824

Truong, X.-T., and Ngo, T.-D. (2017). "To Approach Humans?": A Unified Framework for Approaching Pose Prediction and Socially Aware Robot Navigation. *IEEE Trans. Cogn. Developmental Syst.* 10, 557–572. doi:10.1007/s11370-017-0232-y

Tsoi, N., Hussein, M., Espinoza, J., Ruiz, X., and Vázquez, M. (2020). "Sean: Social Environment for Autonomous Navigation," in Proceedings of the 8th International Conference on Human-Agent Interaction, 281–283.

Uteshev, A. Y., and Goncharova, M. V. (2018). Point-to-Ellipse and Point-to-Ellipsoid Distance Equation Analysis. *J. Comput. Appl. Math.* 328, 232–251. doi:10.1016/j.cam.2017.07.021

Uteshev, A. Y., and Yashina, M. V. (2015). Metric Problems for Quadrics in Multidimensional Space. *J. Symbolic Comput.* 68, 287–315. doi:10.1016/j.jsc.2014.09.021

Vascon, S., Mequanint, E. Z., Cristani, M., Hung, H., Pelillo, M., and Murino, V. (2014). "A Game-Theoretic Probabilistic Approach for Detecting Conversational Groups," in *Asian Conference on Computer Vision* (Springer), 658–675.

Vázquez, M. (2017). *Reasoning about Spatial Patterns of Human Behavior During Group Conversations with Robots. Ph.D. Thesis*. Carnegie Mellon University.

Vázquez, M., Carter, E. J., McDorman, B., Forlizzi, J., Steinfeld, A., and Hudson, S. E. (2017). "Towards Robot Autonomy in Group Conversations: Understanding the Effects of Body Orientation and Gaze," in 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (IEEE), 42–52.

Vázquez, M., Carter, E. J., Vaz, J. A., Forlizzi, J., Steinfeld, A., and Hudson, S. E. (2015a). "Social Group Interactions in a Role-Playing Game," in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, 9–10.

Vázquez, M., Steinfeld, A., and Hudson, S. E. (2015b). "Parallel Detection of Conversational Groups of Free-Standing People and Tracking of Their Lower-Body Orientation," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 3010–3017.

Vázquez, M., Steinfeld, A., and Hudson, S. E. (2016). "Maintaining Awareness of the Focus of Attention of a Conversation: A Robot-Centric Reinforcement Learning Approach," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 36–43.

Yang, F., and Peters, C. (2019). "Appgan: Generative Adversarial Networks for Generating Robot Approach Behaviors Into Small Groups of People," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1–8. doi:10.1109/ro-man46459.2019.8956425

Yang, F., Yin, W., Björkman, M., and Peters, C. (2020a). "Impact of Trajectory Generation Methods on Viewer Perception of Robot Approaching Group Behaviors," in 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 509–516. doi:10.1109/ro-man47096.2020.9223584

Yang, F., Yin, W., Inamura, T., Björkman, M., and Peters, C. (2020b). "Group Behavior Recognition Using Attention-And Graph-Based Neural Networks," in Proceedings of the 24th European Conference on Artificial Intelligence.

Yang, S.-A., Gamborino, E., Yang, C.-T., and Fu, L.-C. (2017). "A Study on the Social Acceptance of a Robot in a Multi-Human Interaction Using an F-Formation Based Motion Model," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 2766–2771. doi:10.1109/iros.2017.8206105

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R., and Smola, A. (2017). Deep Sets. arXiv:1703.06114.

Zen, G., Lepri, B., Ricci, E., and Lanz, O. (2010). "Space Speaks: Towards Socially and Personality Aware Visual Surveillance," in Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis, 37–42.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

REVIEW
published: 27 January 2022
doi: 10.3389/frobt.2021.720319

Helping People Through Space and Time: Assistance as a Perspective on Human-Robot Interaction

Benjamin A. Newman*, Reuben M. Aronson, Kris Kitani and Henny Admoni

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, United States

As assistive robotics has expanded to many task domains, comparing assistive strategies among the varieties of research becomes increasingly difficult. To begin to unify the disparate domains into a more general theory of assistance, we present a definition of assistance, a survey of existing work, and three key design axes that occur in many domains and benefit from the examination of assistance as a whole. We first define an assistance perspective that focuses on understanding a robot that is in control of its actions but subordinate to a user's goals. Next, we use this perspective to explore design axes that arise from the problem of assistance more generally and explore how these axes have comparable trade-offs across many domains. We investigate how the assistive robot handles other people in the interaction, how the robot design can operate in a variety of action spaces to enact similar goals, and how assistive robots can vary the timing of their actions relative to the user's behavior. While these axes are by no means comprehensive, we propose them as useful tools for unifying assistance research across domains and as examples of how taking a broader perspective on assistance enables more cross-domain theorizing about assistance.

Keywords: human robot interaction, assistive robotics, socially assistive robotics, physically assistive robotics, collaborative robotics, rehabilitative robotics

Edited by:
Séverin Lemaignan, Pal Robotics S. L., Spain

Reviewed by:
Cristina Urdiales, University of Malaga, Spain
Yomna Abdelrahman, Munich University of the Federal Armed Forces, Germany

****Correspondence:***
Benjamin A. Newman
newmanba@cmu.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *04 June 2021*
Accepted: *25 November 2021*
Published: *27 January 2022*

Citation:
Newman BA, Aronson RM, Kitani K and Admoni H (2022) Helping People Through Space and Time: Assistance as a Perspective on Human-Robot Interaction. Front. Robot. AI 8:720319. doi: 10.3389/frobt.2021.720319

1 INTRODUCTION

Smart wheelchairs navigating easily through crowded rooms, coaching robots guiding older adults through stroke rehabilitation exercises, robotic arms aiding motor-impaired individuals to eat a meal at a restaurant: these are all examples of research in areas as disparate as intelligent motion planning, rehabilitative medicine, and robotic manipulation that have been independently identified as being able to contribute to the development of robots that can do helpful things for people. This research has been fruitful, but has remained siloed as researchers from these various fields focus on the specific assistive tasks relevant to their own disciplines.

A lack of common structure in the field of assistive robotics makes it difficult for researchers to incorporate findings from other domains into their own work. For example, how does the relationship between a grocery stocking robot and the surrounding customers relate to the relationship between an airport guide robot and the surrounding crowd? Does a robot designed to autonomously declutter a room convey a similar sense of agency as a virtual robot suggesting an optimal ordering in which you should clean your room? Answers to these and similar questions would form a basis that would provide clarity for research in assistive robotics, but are currently difficult to determine due to the disparate nature of assistive robotics.

In this work, we identify a subset of common challenges and develop themes that begin a conversation about how assistance abstracted from specific problem domains and can be used to answer questions about assistance generally, thereby benefiting the entire field of assistive robotics. This would enable researchers to explore the underlying principles of assistive robotics and communicate them across domains. To start, we suggest that assistance is not a characteristic of a robotic system as it has been historically treated. Instead, assistance is a task-independent perspective on human robot interaction. Treating assistance as a task-independent perspective on HRI, we can group existing assistive research by its effect on three key axes: people (e.g., who is involved in the system and the roles they play), space (e.g., how the robot's action affects the task), and time (e.g., when the robot performs its actions during the task).

This perspective considers an assistive system as an interaction in which a user and a robot forge a complex, asymmetric relationship guided by the user's goals. This perspective is somewhat different from general HRI because the user is responsible for determining the interaction's end goal while the robot acts in service of this goal. Similar to other collaborative settings, the human-robot pair is then tasked with performing subsequent actions to achieve the human's goal, but unlike some collaborations, maintaining human autonomy is paramount. In this relationship, the robot has more agency and independence of action choice than a simple tool (i.e., the robot's choice of action is not determined solely by the user), but it must defer to the user's goal and independent actions.

We introduce three design dimensions with which roboticists can begin to reason about the assistive interactions of robots and humans. First, we discuss how the assistive robot's role can be described with respect to the relationship it has with its user, for example, how it weighs priorities when there are multiple potential people it could assist. Second, we propose that an assistive robot's role can be described in terms of how it operates in the execution space, that is, the space in which the robot has its primary effect. Finally, we propose that the same robot's actions can be described in terms of the temporal space, that is, the duration and sequence of the actions. We support these dimensions by reviewing and grouping over 200 recent assistive robotics research papers.

By using assistance as a lens through which to analyze patterns that arise in assistive robotics, we hope to help designers of assistive robots more easily explore the design space and identify similar examples of past solutions, even across application domains. Additionally, we hope this work will motivate researchers to continue to refine this notion of assistance and its effects on human-robot interaction paradigms.

2 THE ASSISTANCE PERSPECTIVE

In the field of robotics, defining assistance can be tricky. In a broad sense, every robot is built to assist some person. Therefore, we do not attempt to separate assistive systems from non-assistive systems. Instead, we propose assistance as a particular perspective through which many robotic systems can be viewed. This perspective considers robotic agents that are autonomous in action but subordinate in goal to a human partner. Almost any robot system can, in theory, be viewed as assistive to someone, so we do not limit this scope. Rather, we explore what this analytic framework provides. This perspective clarifies particular design tradeoffs and trends general to assistive systems whatever their task domain. In this work, we describe several key design axes that arise when considering a robotic system as assistive and discuss implications these axes have on the interaction.

Before discussing these key design axes, we first formalize what we mean by a human-robot interaction, then provide a more detailed description of what it means to view assistance as a perspective. Next, we give a brief synopsis of previous attempts to characterize assistance and assistive robotics, and finally we give an overview of the remainder of this paper.

2.1 General Human-Robot Interaction

Before discussing assistance, we first sketch a general framework for human-robot interaction, which we draw broadly from multi-agent systems research. Formalizations of this problem can be found in previous literature (Jarrassé et al., 2012); here we only establish enough language to discuss assistance rather than requiring assistive systems to use this exact model.

First, we define a user $u \in U$ as any person involved closely in the interaction. Typically, the user is in close physical proximity to the robot and provides explicit or implicit control signals to the robot. For example, a person teleoperating a robotic arm, getting directions from a social robot, or building a table with a robot helper, would be considered a user.

Next, the system has at least one robot $r \in R$. Canonically, a robot is defined as an embodied system that can sense its environment, plan in response to those sensory inputs, and act on its environment. An assistive robot may have a wide array of sensory, planning, and acting capabilities in order to be successful in its task. Some of these capabilities will be critical for the robot's functioning (e.g., LIDAR to avoid hitting obstacles), while others will be critical for providing assistance to the user (e.g., a body pose recognition algorithm to identify the user's location and gestures).

Finally these agents exist in a shared environment, each with its own internal state. These are described in totality by the mutual state $s_m = (s_r, s_u, s_e)$ that defines the individual states of the robot, user, and environment. The robot and user both have goals $g_r, g_u \in G$ and can take actions $a_r \in A_r$ and $a_u \in A_u$ that affect their mutual state. By acting to update their mutual state, each agent has the potential to affect the other agent's behavior resulting in an interaction between the two agents. Depending on the exact scenario, a task will be considered complete when one or more agents has achieved their goal.

2.2 Assistance as a Perspective on Human-Robot Interaction

Using this formulation, we can more carefully define assistance. Assistive systems interpret the robot as autonomous in its actions but subordinate in its goal. By giving the user the sole responsibility for setting both agents' goals, the two agents now attempt to satisfy some shared goal g by reaching a mutual state where g is true: s_m^g. This framing distinguishes assistive robotics from both traditional assistive technologies like a white cane, which has no control over its actions or goals, and traditional robotics, which develops systems with full control over their actions and goals. This framing gives rise to three key design axes: how assistive robots affect *people* through *space* and *time*. The discussion of these implications is the subject of the rest of this paper.

In HRI, as in assistive robotics, there is no requirement for there to be a single user. In fact, many assistive robotics scenarios involve more than one user. This becomes challenging, as it is the responsibility of one of these users to set the goal for the robot, but selecting which user has this responsibility may change the type of assistance the robot is able to provide. This is especially true when one user's goals may conflict with another user's goals. This highlights the importance of determining the roles of people when considering assistive robotics problems (**Section 4**).

Furthermore, since the user and robot are working to accomplish the same goal, the robot has freedom over its action space. As a baseline, the robot can assume the user would perform the task independently, without its aid. The robot can then choose its action space to align with how it can most beneficially assist the user over this baseline scenario. In addition to the standard strategy of directly manipulating the environment, the robot can assist by altering the user's state space, encouraging the user to make more effective task progress. For example, a head-mounted augmented reality device displaying the optimal path for cleaning a room can assist the user without needing to physically interact with objects. Assistive scenarios allow more choice over the robot's action space than would a general robot (**Section 5**).

Finally, in order to advance to the mutual goal state and complete the task, the user and robot each complete a sequence of actions ($a_u^1, \ldots, a_u^t, a_r^1, \ldots, a_r^t$, respectively) that transition the system to the desired goal state ($s_m = s_m^g$). Given that these actions occur in the mutual state, it is important that the user and the robot time their actions appropriately, so that they do not attempt to solve the same part of the task simultaneously, or worse, provide conflicting actions that result in undoing each other's work. How to time actions is crucial to studying assistive robotics (**Section 6**).

Each of these axes presents researchers with decisions that result in critical trade-offs when designing an assistive robot. Throughout the remainder of this work, we will describe how assistive robots from different application domains fall along these axes.

By taking assistance as a perspective, it is our goal to provide an abstraction that allows for comparing systems from different domains to discover universal challenges that arise from robot assistance. We do not suggest that these axes describe a full assistive system or are a complete set of critical design axes. Rather, viewing assistance along these particular axes of people, space, and time enables some cross-domain comparisons and insights on its own, and it also demonstrates how assistance overall can benefit from a general examination.

2.3 Prior Categorizations of Assistive Robotics

By grouping assistive robots along the aforementioned design axes, we view assistance as an abstract concept that illuminates parallel research problems across different application domains. We build on previous literature which categorizes assistive robotics within particular application domains, for example socially assistive robots (Fong et al., 2003; Matarić and Scassellati, 2016), joint action (Iqbal and Riek, 2019) and physically assistive robots (Brose et al., 2010).

Some work does try to describe assistance as a whole. Jarrassé et al. (2012) categorizes joint action between dyads by positing a cost function for each agent defined on each agent's task error and required energy. Among categories in which both agents are working together towards the same goal, the paper specifies collaboration between two equal peers, assistance when one agent is subordinate to another, and education in which the educator assists the partner but moderates its own effort to encourage increasing effort from its partner. We take this core idea of assistance as subordination and build on it in our definition of the assistance perspective.

Most similar to the current work, perhaps, is the accounting given in Wandke (2005). This overview of assistance in human-computer interaction notes that defining assistance as any system that provides some benefit to the user would include nearly all technical artifacts. Therefore, the paper restricts its attention to systems that bridge the gap between a user and the technical capabilities of the system due to the user's unfamiliarity with the system or excessive burden of use. In contrast to this approach, our work presents assistance as a perspective rather than a definition; it could in principle be applied to any technical artifact but may only be useful for some. Additionally, this definition of assistance focuses on how assistive systems correct a deficiency in a user's understanding of the system or capability to use it. In contrast, our definition of assistance as a perspective admits beneficial actions from the robot of all sorts, not just those repairing the user's ability to use a system.

2.4 Overview of This Paper

By defining assistance as a perspective, we provide language to discuss ideas about assistance from different domains. This will allow researchers from various areas of assistive robotics to come together to illuminate and discuss common research challenges. Additionally, researchers can make design decisions about how the assistive robot affects people in space and time by using this framework to consider similar approaches to problems from disparate task domains. In the remainder of this paper, we discuss these design axes and explore their implications through a review of existing assistive robotics literature. **Section 3** describes our method for collecting these papers **Section 4** describes the people design axis, **Section 5** describes the space design axis, and **Section 6** describes the time design axis. These axes are summarized in **Table 1**. We then conclude the paper with a discussion over the implications of this work.

3 METHODS

To develop this taxonomy, we conducted a literature review of recent papers on assistive robotics.

TABLE 1 | Assistive robots can be explored along three key axes: how the assistive system thinks about additional people, what part of the mutual state aligns with its action space, and at what time it executes its actions during a task.

Key axis	Description
People (**Section 4**)	How the robot considers additional people outside the baseline dyad.
Targets of assistance	Additional people whose goals are of comparable importance to the user.
Interactants	Additional people whose goals are not privileged and use general human-robot interaction approaches.
Space (**Section 5**)	The portion of the mutual state the robot's actions affect.
Environment	The robot affects the environment directly by, e.g., manipulating task objects.
Human body	The robot affects the user's body by physically moving some portion of their body.
Human brain	The robot affects the user's mental state by providing information about the task or reducing the cognitive burden.
Time (**Section 6**)	The relative timing between a robot's actions and the user's explicit commands during the task.
Proactive	The robot acts before an explicit command.
Reactive	The robot acts in response to an explicit command.
Simultaneous	The robot acts simultaneously with user action.

3.1 Initial Search

First, we hand-selected 74 papers from the last 5 years of the annual Human Robot Interaction conference (HRI 2016–2020). From these papers we generated an initial set of search terms by aggregating titles, abstracts, and author generated keywords using the R (R Core Team, 2017) package litsearchr (Grames et al., 2019). Using these aggregated keywords, we formed an initial search query.

3.2 Refined Search

We ran the initial search query on the Web of Science. This search yielded approximately 1,500 papers. We repeated the keyword aggregation on this set of keywords, and then hand-selected new keywords from among them based on their prevalence and relevance to assistive robotics. We repeated the Web of Science query with this refined set of keywords, which yielded, again, approximately 1,500 papers. The refined search was run on 29th January 2021. We included a paper based on whether the following statement evaluated true based on a search of the entire text of the paper.

((assist* NEAR *robot*)

OR (collab* NEAR *robot*))

AND (*human* OR *people* OR *person* OR *subject* OR *user* OR "elderly people" OR "older adults" OR "natural human" OR "stroke patients" OR "healthy subjects")

AND ("human-robot interaction" OR "human-robot collaboration" OR "robot interaction" OR "robot collaboration" OR collaboration OR hri OR "human robot collaboration" OR "physical human-robot interaction" OR "human robot interaction" OR "machine interaction" OR "human-machine interaction" OR "human interaction")

AND ("collaborat* task*" OR "assembly task*" OR "social interaction*" OR "assembly process*" OR "shared workspace*" OR "manipulation task*" OR "human safety" OR "daily living" OR "service *robot*" OR "production system*" OR "safety standard*" OR "mobile robot*" OR "assisted therap*" OR "collision avoidance" OR "object manipulation" OR "collaborative assembly" OR "socially assistive" OR "assistive *robot*" OR "social *robot*" OR "teleoperat*"))

3.3 Paper Selection

Starting from the refined Web of Science results, we filtered out all papers from venues with fewer than two related documents and papers that were older than 5 years, with a small exception. In an attempt to keep papers with significant contributions to the field, papers older than 5 years were kept if they had more than 10 citations. This process left approximately 465 papers. Each paper in this set was then manually checked for relevance by reading the title and abstract. To be included, we required the paper to include both 1) an assistive interaction with the user and 2) a system capable of taking actions. This step mainly removed papers focused on robotic system development or perception improvements rather than assistance itself. This yielded 313 papers, each of which was again reviewed against the aforementioned exclusion criteria. The entire search process yielded over 200 papers that we classified into our taxonomy.

4 PEOPLE

In **Section 2**, we described assistance with single users. This description works well for situations that have only one user, which is common in laboratory settings. In realistic settings, however, a robot will typically encounter more than one person in the course of completing their task. These other people can act in a variety of different roles within the interaction. In this section, we explore themes in how assistive interactions incorporate more people into the general human-robot dyad (**Figure 1**).

4.1 Terminology

The simplest approach a system can take towards other people is simply to ignore them completely. While this case tends not to be analyzed explicitly, it is implicit in many systems. This strategy can be appropriate, especially during situations in which encountering additional people is rare. When working with other people, though, the robot could implicitly account for additional people by relying on its primary user to provide

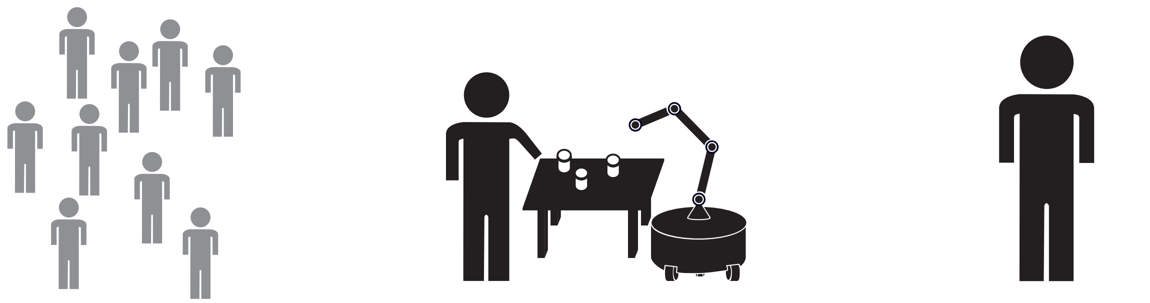


FIGURE 1 | An assistive system can treat people beyond a single user as additional targets of assistance or as interactants, and either choice introduces particular complications into the assistive dynamic.

controls that appropriately consider other users. Finally, a robot might intentionally downplay its relationship to additional people when accounting for them would conflict with its primary user's goals, such as an emergency response robot that ignores standard social navigation behaviors to reach its patient as fast as possible.

When the system does choose to reason about other people, its treatment of them can be determined by dividing them into two different roles: the target of assistance, whose goals are of equivalent importance as other targets; and interactants, who require the attention owed to any other person as explored throughout human-robot interaction research but don't have their goals privileged by the robot.

A target of assistance derives directly from the definition of assistance: an assistive scenario must support the goals of at least one person. Consider a scenario in which a person who has a spinal cord injury uses a robotic arm to aid them in eating a meal with friends at a restaurant. In this scenario, the arm's user sets the goal for the robot: to bring food from their plate to their mouth so they can consume it.

The second role a person can play in an interaction is that of interactant. An interactant is any other person involved in the scenario who is not a target. Continuing the previous example, the people who are out to dinner with their robot-operating friend are interactants. They have no direct bearing on the robot's goal, but they are potentially affected by the robot's actions and may require some design effort for the system. For example, the robot may have to avoid collisions with them during its operation. While the robot's relationship to interactants is not assistive, the presence of a specific target of assistance can affect how the robot interacts with others.

When considering assistive systems that involve more than a single target, the system must determine in which of these roles to consider the additional people. These two roles are not mutually exclusive; there can be more than one of each in a given scenario. Additionally, both targets of assistance and interactants can give explicit control input to the robot. Designating people as additional targets or as interactants brings about different challenges for the assistive system.

4.2 Additional Targets of Assistance

One challenge arising from a single robot having multiple targets of assistance is that the goals issued by these targets can conflict with one another. In the eating scenario, the robot might instead be assisting everyone present, perhaps by both feeding its user and serving food to other people at the table. Here, the robot is presented with a conflict: how should it choose to prioritize the goals given by its targets and reconcile differences between them?

This can be especially challenging in contexts such as education. An educational robot might consider the teacher as its target and work to enrich a student according to a mandated curriculum. It can also consider the student as its target and try to engage the student with concepts that are interesting to them regardless of the curriculum. Much research in this area aims to make the content proposed by the teacher more enjoyable by developing robotic behaviors that are meant to keep the student engaged. Leite et al. (2015) designed a robot puppet show to engage young learners in an educational story, Martelaro et al. (2016) designed a robot that encourages students to develop trust and companionship with their tutor, and Christodoulou et al. (2020) designed a robot to give nonverbal feedback to students in response to quiz answers to keep them engaged with the testing material. In contrast, Davison et al. (2020) took a different approach and developed the KASPAR robot to look like another student and deployed it in unsupervised interactions that were totally motivated by the student. In this way, they allowed the student to approach the learning material voluntarily, giving the student more agency to learn what they desired and at their own pace.

This dilemma can again be seen in therapeutic contexts, where a robot must reconcile the goals of the doctor and the patient. Robots can increase a patient's motivation to do mundane, repetitive or uncomfortable exercises through the use of a robot that does the exercise alongside the patient (Tapus et al., 2007; Schneider and Kummert, 2016). Alternatively, a robot could be used to give the patient more agency and independence over their own treatment by helping someone independently practice meditation (Alimardani et al., 2020), do independent cognitive behavioral therapy (Dino et al., 2019), or home therapy for autism (Shayan et al., 2016).

A full analysis of these interactions treats both the teacher and the student, or both the therapist and the patient, as targets of assistance with goals that often align but are not identical. This alignment mismatch can often lead to ethical challenges, which are even more fraught when the capabilities, agency, and relative power of the possible targets vary. While there is no general technical solution, this language encourages designers to explicitly enumerate the multiple targets of the assistance and to reason directly about conflicts in their goals.

4.3 Additional Interactants

On the other end of the spectrum are robots that treat additional people in the system as interactants. Robots designed with this relationship in mind prioritize the goals of its target of assistance. In our assisted eating scenario, the robot may need to follow basic social norms around the other diners by avoiding collisions with them, but it does not privilege their goals.

This relationship is typically used in scenarios where some figure of authority (e.g., a teacher or a therapist) needs to relieve themselves of some amount of work. For example, a teacher could employ a robot to teach half of their class in order to reduce the student-to-teacher ratio for a particular lesson (Rosenberg-Kima et al., 2019), or even have the robot teach the class alone if they need to finish other work (Polishuk and Verner, 2018). In this way, the teacher is the target of assistance, while the students are treated only as interactants. The robot should be able to teach competently enough to achieve the teacher's goals, but the students' preferences about using the robot are not of direct concern.

Similarly in emotional or physical therapy a robot can be employed to lead group sessions in lieu of a doctor, who may have more classes than they can handle (Fan et al., 2016; Ivanova et al., 2017). Alternatively, the robot may be better at collecting certain information than the user. For example a patient who has suffered a stroke may be unable to emit certain social signals expected during social interaction. This could negatively affect a doctor's opinion of this patient, a problem that could be circumvented by having a robot collect this information (Briggs et al., 2015; Varrasi et al., 2019). The patient here, however, is not asked whether they may prefer the social interaction regardless of the implicit bias the doctor may possess.

These systems don't generally follow an assistance dynamic with interactants, rather, general human-robot interaction research applies. However, the fact that the system has a target, even if the target is not present, can change the robot's behavior: a robot acting as a proxy for a specific teacher may have different behavior than one employed as a general-purpose robot, which might have bearing on how the general human-robot interaction problem is resolved.

4.4 Combinations of Roles

If an assistive robot has multiple additional people present in the interaction, it can choose to consider some of them as targets and others as interactants. In this relationship, our assisted eating robot might treat both the user and the companion seated next to them as targets of assistance, while those eating companions seated further away from the user are treated as interactants. In this way the robot can carefully maintain the goals of multiple people in proximity to the robot. This framework can allow for more complex robot behavior near to the user without the additional complication of handling everyone else at the table.

Another example would be a robot that participates in a collaborative scenario with multiple human actors, some of whom serve as both targets of assistance and interactants, while others are only interactants. For example, consider a local repair-person who needs help from a remote repair person. To give instructions, the remote repair person can use a robot to highlight the parts of the environment they are discussing (Machino et al., 2006). In this way, both actors are interactants in the scenario, but only the local repair person is the target of assistance.

4.5 Implications

These various relationships clarify the design choices involved in developing an assistive system. A particular task, such as assistive eating, does not require a particular relationship between the robot and the people it encounters. Rather, how a robot relates to these people is a design decision that will have implications as to how the task is completed.

The choice of roles affects how assistive systems with multiple people are evaluated. When treating the user and their eating companions all as targets of assistance, the robot would need to verify that it is helping them all in achieving their independent goals. This type of evaluation may be difficult to actually measure and nearly impossible to succeed on, as the companions have conflicting interests from the user. Identifying what type of relationship the robot should have with its users can help researchers disambiguate otherwise similar systems to determine which evaluations are important.

The choice of which roles to use may also have implications on how much autonomy to imbue in the robot. A robot that balances the goals of many people may require complex sensing, modeling, and planning to carefully moderate between them. A simpler robot might delegate this goal moderation problem to its user and treat additional people as interactants or ignore them entirely. This system gives the target more control over the goals, but requires additional input from the user. If the robot maintains full autonomy in this scenario, but it does not plan for other people's goals, it may in fact endanger them by running into them where another system would have chosen to avoid them. These ideas show how the choice of relationship between the robot and the people it encounters throughout a task can impact the design of the final system.

5 SPACE

Assistive robotic systems can perform similar tasks by acting in different action spaces. We show in **Section 2** how to represent the mutual state during the interaction as the state of the user s_u, the state of the robot s_r, and the state of the environment s_e. In general, a user employing an assistive robots is aiming to make some alteration to s_e. Since the robot is tasked with aiding the user and not directly accomplishing this state alteration, the robot can assist the user by making a change to any part of the mutual state

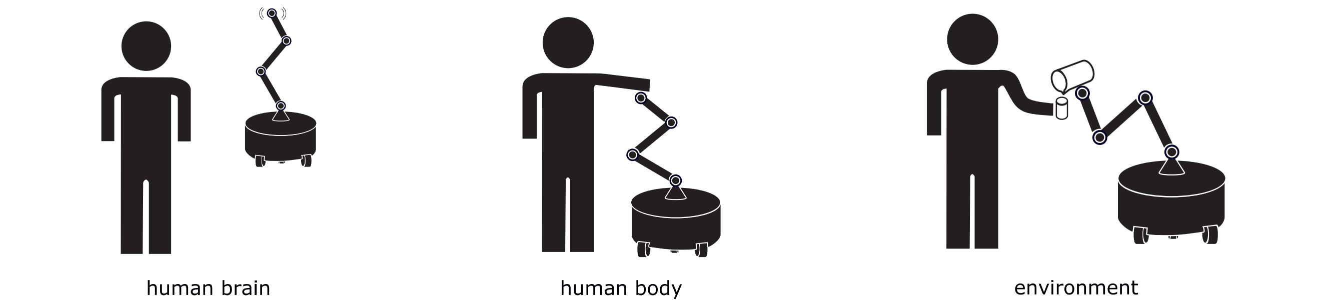


FIGURE 2 | A robot can provide assistance by acting in several different action spaces. It can assist by giving information to the user, adjusting the user's body, or changing the environment to help complete the task.

that makes it easier for the user to accomplish their goal. In this maner, a robot can provide many different types of assistance when helping to complete the same overall task.

Consider an assistive eating robot. The robot and its user sit at a table across from one another, with a plate of food between them. The user's goal is to eat the food. The robot can provide assistance by performing a variety of different actions: it can act on the user's mental state by projecting a light onto a morsel of food that would be easy to grab next, it can change the physical state of the user by guiding their hand into an appropriate position, or it can change the environment by picking up the morsel and feeding it to the user. All of these action spaces apply to the same task and the same goal; what differs is in what way the user would most benefit from assistance.

To illustrate this point more broadly, we provide a review of recent assistive robotics literature, grouped by whether the robot is acting on the user's mind, user's body, or environment (**Figure 2**).

5.1 Environment

One straightforward assistive robot is one that simply completes a task for the user. For example, research has focused on autonomous butler robots (Srinivasa et al., 2010, 2012) that perform tasks such as cooking and cleaning. Such a robot assists a user by navigating around the apartment picking up misplaced items such as dirty laundry and dishes and placing them in appropriate locations such as a laundry hamper or dishwasher. The robot provides assistance by directly changing the environment. To meet the minimal requirement of providing assistance (i.e., delivering some benefit to the target of assistance), the robot must shift the environment from an undesirable state configuration to a more desirable one.

Much research surveyed here assists users in exactly this way: by providing autonomous assistance through environmental state manipulations. Researchers have explored how a user can command a robot to organize a messy room (Mertens et al., 2011; Cremer et al., 2016; Koskinopoulou et al., 2016; Pripfl et al., 2016; Jensen et al., 2017), fetch misplaced or distant items (Iossifidis and Schoner, 2004; Unhelkar et al., 2014; Huang and Mutlu, 2016; Wieser et al., 2016), or even perform more specialized tasks autonomously (under the direction of the user) such as assisted eating (Canal et al., 2016) and other tasks of daily living (Nguyen and Kemp, 2008), search and rescue (Doroodgar et al., 2010), welding (Andersen et al., 2016a), or other industrial tasks (Mueller et al., 2017). Assistive tasks performed autonomously at the request of a user through environmental manipulation can provide several benefits. This method of task execution requires little user input, which makes it efficient for users who prefer not to spend time on chores and beneficial for users who may not be able to accomplish the task at all.

Environmental assistance is not solely the domain of autonomous robots, however. Collaborative robots, specifically in tasks where the user and the robot take independent actions that jointly manipulate the environment towards a mutual goal state, also perform environmental assistance. Examples of such systems include collaborative cleaning (Devin and Alami, 2016) and assembly (Savur et al., 2019; Zhao et al., 2020). A robot working collaboratively with a user can improve its efficiency by modeling the user's behavior, for example by determining specific poses to hold an object in to facilitate fluid collaboration during assembly (Akkaladevi et al., 2016) or by anticipating and delivering the next required item in assembly (Hawkins et al., 2013, 2014; Maeda et al., 2014) or cooking (Koppula et al., 2016; Milliez et al., 2016), or by providing help under different initiative paradigms during assembly (Baraglia et al., 2016). Collaborative environmental assistance can also be used to perform joint actions with a user, such as in handovers (Cakmak et al., 2011; Kwon and Suh, 2012; Grigore et al., 2013; Broehl et al., 2016; Canal et al., 2018; Cserteg et al., 2018; Goldau et al., 2019; Lambrecht and Nimpsch, 2019; Nemlekar et al., 2019; Newman et al., 2020; Racca et al., 2020), where the goal is to transfer an object from the robot's end effector to the user's hand; or co-manipulation (Koustoumpardis et al., 2016; Nikolaidis et al., 2016; Schmidtler and Bengler, 2016; Schmidtler et al., 2016; El Makrini et al., 2017; Goeruer et al., 2018; Rahman, 2019b; DelPreto and Rus, 2019; Rahman, 2020; Wang et al., 2020), where the aim is for the user and the robot to jointly move an object to a specified location or provide redundancy in holding an object in a joint assembly task (Parlitz et al., 2008) or safety critical situation such as surgery (Su et al., 2018).

So far, all examples of environmental assistance have been provided by standalone robots, commonly taking on a humanoid or robotic arm morphology. These robots affect the environment by changing their own configurations first (e.g., using a robot arm to pick up an object). As such, they are considered decoupled from the environment. Robots can also be designed to be coupled with the environment; in these examples, it is hard to distinguish between the robot's state and the environment state. These robots often take on more conspicuous yet specialized morphologies, such as a mechanical ottoman (Sirkin et al., 2015; Zhang et al., 2018). For example, a robotic suitcase can assist an airline passenger by following them through an airport (Ferreira et al., 2016) and manipulating the user's sense of trust by moving across various proxemic boundaries. A set of robotic drawers containing tools can assist a user in completing an assembly by proactively opening the drawer containing the next required tool (Mok, 2016), and it can also manipulate a user's enjoyment in completing the task by employing emotional drawer opening strategies. Environmentally coupled robots can be designed to be "invisible," (Sirkin et al., 2015) or to be modifications to an existing environment or object. Moving away from more traditional robot appearances may mitigate any negative effects from interacting with a robot.

Other approaches include shared control which separates the responsibilities of the user and the robot during the task. For example a teleoperated surgery robot can hold a patient's skin taut so that the surgeon can focus on performing incisions (Shamaei et al., 2015). A telepresence robot (Kratz and Ferriera, 2016) can automatically avoid obstacles during navigation (Acharya et al., 2018; Stoll et al., 2018) or automatically rotate its camera to keep a desired object within view (Miura et al., 2016). Finally, a remote, teleoperated space robot can perform as much of a task as is possible before it pings the space station for human intervention (Farrell et al., 2017). By having the robot configure itself according to some of the task requirements, the robot allows the user to focus on other parts of the task.

5.2 Human Body

While assistance applied directly to the environment can solve a wide variety of tasks, some tasks require alternate strategies. One such scenario is when some change to the user's physical state is required to perform the task. For example, consider a robot designed to assist a user who has difficulty bathing themselves. While it is technically possible for that robot to transform the environment by bringing a bathtub to the user, this is obviously impractical. The robot can instead transform the user's state by bringing them closer to the bathtub (Dometios et al., 2017; Papageorgiou et al., 2019). This strategy of moving a user to assist them is similar to autonomous environmental manipulation, but now the user is being manipulated instead of the environment. This strategy results in limited agency to the user, and is typically only employed when the user has minimal ability to complete the task themselves.

In cases where users can perform some aspects of the task, a robot can also assist by supplementing a user's existing abilities. For example, if a user can walk but has difficulty balancing or navigating, a smart walker can be utilized to help the user navigate between locations (Papageorgiou et al., 2019; Sierra et al., 2019). Similarly, if a user has some control over their limbs, an exoskeleton robot can be used to provide extra support for day-to-day usage (Baklouti et al., 2008; Lim et al., 2015; Choi et al., 2018; Nabipour and Moosavian, 2018) or in therapeutic scenarios in order to help a user strengthen weakened muscles (Carmichael and Liu, 2013; Zignoli et al., 2019).

In addition to aiding in task execution, physical user state manipulation can also be used to assist in planning, such as when a user's sensing capabilities are diminished. For example, a visually impaired user may wish to solve a Tangram puzzle but must pick up and feel each piece individually. To provide assistance to the user, a robot could sense the puzzle pieces and determine which pieces are viable for the next step of assembly. The robot can then physically guide the user's hand to this piece allowing the user to solve the puzzle (Bonani et al., 2018). This is an example of human body state manipulation. Instead of manipulating the environment to solve the task, the robot instead changes the user's physical state configuration in order to better position them to solve the task.

Robot assistance that acts on a user's body can also be done by using the resistance of the robot's own joints. A user kinesthetically manipulating a robot arm, for example, may not know the exact path the arm should travel in order to complete a co-manipulation task. The robot can change its admittance or transparency such that it becomes easier (Jarrasse et al., 2008; Li et al., 2015; Lee and Hogan, 2016; Mariotti et al., 2019; Muthusamy et al., 2019; Luo et al., 2020) or more difficult (Bo et al., 2016; Kyrkjebo et al., 2018; Cacace et al., 2019a,b; Wu et al., 2020) to move as the robot's end effector deviates from a known, low-cost path. This idea can also be applied to full-scale robots, allowing a user to navigate a robot from one point to another by guiding it as if it were another human (Chen and Kemp, 2010) or to use the stiffness of the robot's arm as a support while standing up (Itadera et al., 2019). Admittance control as a body state manipulation allows the user to have a high degree of control when operating the robot, but allows the robot to provide information about which parts of the environment are better to traverse by altering the stiffness of its joints. This strategy can also be used in therapeutic settings, where a patient recovering from a stroke can be given an automatic, smooth schedule of rehabilitation exercises as the robot changes its admittance depending on the force feedback it receives from the user (Ivanova et al., 2017).

5.3 Human Brain

The final location of assistance we identify is the user's mental state. These robots assist by transforming the user's understanding of the world in a helpful way. One common method is for the robot to communicate unknown environmental information to the user. For example, a robot can play particular sounds as it completes its tasks so that a user can track it more easily (Cha et al., 2018). A robot can also describe the local environment for a visually impaired user in a navigation task, enabling them to create a semantic map of the environment (Chen et al., 2016). Similarly, a robot can provide a visual signal to designate objects it intends to interact with so the user can avoid them (Machino et al., 2006; Andersen et al., 2016b; Shu et al., 2018), areas where the robot expects to move so the user can stay away (Hietanen et al., 2019), or areas or paths that the robot thinks the user should take to complete a task in an optimal fashion (Newman et al., 2020). In an emergency scenario, a robot can visually indicate the direction of a safe exit (Robinette et al., 2016). Finally, a robot can provide haptic feedback to indicate when to turn in a navigation task (Moon et al., 2018; Li and Hollis, 2019). Robots that provide alerts like these assist by communicating information about the task or the environment directly to the user so that the user can effectively perform the task.

Robots can also assist in the mental state domain by adopting social roles. Generally, these robots are designed to perform socially beneficial functions similar to those that a human would provide, such as a robot that takes the role of a customer service agent (Vishwanath et al., 2019) or a bingo game leader (Louie et al., 2014). In educational settings such as one-on-one tutoring (Kennedy et al., 2016; Fuglerud and Solheim, 2018; Kanero et al., 2018; van Minkelen et al., 2020) and classroom teaching (Kennedy et al., 2016; Ramachandran et al., 2016; Westlund et al., 2016; Polishuk and Verner, 2018; Ono et al., 2019; Rosenberg-Kima et al., 2019), a robot can deliver lectures in a similar manner to a human teacher. In therapeutic and medical settings, a robot can administer routine medical surveys (Varrasi et al., 2019) independent of the doctor's social biases (Briggs et al., 2015), provide therapy sessions for routine cognitive behavioral therapy (Dino et al., 2019) or physical therapy (Meyer and Fricke, 2017), and perform other general therapeutic tasks (Agrigoroaie et al., 2016; Fan et al., 2016; Salichs et al., 2018; Alimardani et al., 2020). Finally, a robot's assistance can vary based on its social role, such as a concierge robot performing different social behaviors when responding to children or adults (Mussakhojayeva et al., 2017), an advice-giving robot providing explanations when a user's behaviors become non-optimal (Gao et al., 2020) or a robot that gives cooking advice varying its strategies so that the advice is more readily received (Torrey et al., 2013).

Instead of performing a procedure itself, a robot can assist a professional when affecting a user's mental state. When a therapist is unable to be physically present with a child, for example, a parrot robot can be employed in the home to entice a child with autism to practice skills learned during a therapy session (Shayan et al., 2016; Bharatharaj et al., 2017). During therapy with agitated patients, introducing a pet-like PARO robot can induce mental states more conducive to effective therapy (Shibata et al., 2001; Sabanovic et al., 2013; Chang and Sabanovic, 2015; Shamsuddin et al., 2017). A child-like robot can allow a young patient to practice social skills with a partner more akin to a peer than the therapist is (Goodrich et al., 2011; Kim et al., 2014; Taheri et al., 2014; Ackovska et al., 2017; Nie et al., 2018). Similarly, a child-like robot can assist a teacher by reinforcing a student's desire to self-engage in educational material, something students may be more likely to learn with a peer than a teacher (Wood et al., 2017; Davison et al., 2020), or increase a user's ability to recall a story by acting out portions of it (Leite et al., 2015).

Since robot actions are sometimes interpreted socially and as being intentional, robots can select their actions to influence the user's mental state. For example, predictable and legible motion strategies that indirectly communicate a robot's goals are readily interpreted by people (Dragan et al., 2013). These same strategies can be used in collaborative tasks to indirectly show the robot's goal to the user (Bodden et al., 2016; Faria et al., 2017; Zhu et al., 2017; Tabrez et al., 2019). Robots can also mimic human nonverbal behaviors like deictic eye gaze and pointing gestures to indicate task-relevant objects during collaborative tasks (Breazeal et al., 2004; Fischer et al., 2015) or to assist in completing mentally taxing tasks (Admoni et al., 2016; Hemminghaus and Kopp, 2017).

Similarly, robots can use their behavior to suggest their internal emotional state. This strategy can increase rapport, fluidity and reception of a robot's assistance through emotive motions (Mok, 2016; Terzioglu et al., 2020) or giving the user feedback regarding a task's success through facial expressions (Reyes et al., 2016; Rahman, 2019a; Christodoulou et al., 2020). Using socially meaningful actions enables assistive robots to communicate with the user efficiently and fluidly.

Robots can also use social behaviors to induce specific, beneficial emotional responses from a user. By mimicking human nonverbal behaviors, robots can use their eye gaze to induce social pressure on a user to work more efficiently (Riether et al., 2012) or to soften its own dominance to allow for better teamwork (Peters et al., 2019). Assistive robotic gestures can also increase feelings of openness in people who are discussing negative experiences (Hoffman et al., 2014) and motivation in users during medical testing (Uluer et al., 2020), in users during physical exercise (Malik et al., 2014; Schneider and Kummert, 2016; Malik et al., 2017), and in stroke patients performing rehabilitative exercises (Tapus et al., 2007). Since people generally view robotic gestures as intentional, robots can use these gestures to induce mental states that assist the user in performing a task.

In addition to nonverbal communication strategies, robots that are capable of speech can converse with users to induce beneficial mental states (Knepper et al., 2017). Robots can use speech to change the content of the conversation (Gamborino and Fu, 2018) or to answer a question about the surrounding environment (Bui and Chong, 2018). Robots can use dialogue to gather information during collaborative teleoperation (Fong et al., 2003), to engender trust in an escape room (Gao et al., 2019), or to facilitate collaboration between two targets of assistance (Strohkorb et al., 2016). Robots can also talk about themselves to influence a user's view of themselves. For example, tutoring robots for children can make vulnerable statements about themselves to increase trust with the student and student engagement (Martelaro et al., 2016). Similarly, a robot in a group setting can facilitate group trust by leading with vulnerable statements about itself, so that its teammates feel more comfortable sharing their own vulnerabilities. This effect can cascade as more group members explain their own failures, console each other, and laugh together (Sebo et al., 2018). Failing to deliver assistance in contexts where the robot is expected to provide assistance can have deleterious effects on a user's mental state, causing users to mistrust the robot and harm their relationship and rapport (Kontogiorgos et al., 2020; Rossi et al., 2020).

Beyond focusing on specific content of speech, conversational robots can further affect the user's mental state in the way they speak. Robots can perform back-channelling to give the appearance of active listening (Birnbaum et al., 2016; Sebo et al., 2020), or give informative feedback to improve task performance (Guneysu and Arnrich, 2017; Law et al., 2017; Sharifara et al., 2018), a user's self-efficacy (Zafari et al., 2019), or their motivation (Mucchiani et al., 2017; Shao et al., 2019). Robots can choose to only interrupt a distracted user at appropriate times (Sirithunge et al., 2018; Unhelkar et al., 2020). A robot can also change its tone to project an emotion such as happiness to improve the user's mood and task performance (Mataric et al., 2009; Lubold et al., 2016; Winkle and Bremner, 2017; Rhim et al., 2019). Finally, a robot can combine these qualities with the content of the conversation to change the user's perception of the robot's social role (Bartl et al., 2016; Bernardo et al., 2016; Monaikul et al., 2020). Specifically, a robot can act as a student during a tutoring session to induce different learning techniques in a human student (Sandygulova et al., 2020).

Shared control, especially when an input controller (e.g., a joystick) limits the number of input degrees of freedom (Aronson et al., 2018), can also be made easier for user's by providing assistance that alters the user's mental state. A robot arm can assist its user by maintaining more easily controllable state configurations (Javdani et al., 2015; Till et al., 2015; Vu et al., 2017; Aronson et al., 2018; Newman et al., 2018) or by optimizing which degrees of freedom the user can control at any given time (Herlant et al., 2016). This idea can be extended to supernumerary arms that provide users with an additional appendage but are difficult to control (Nakabayashi et al., 2018; Vatsal and Hoffman, 2018), teleoperating robotic arms through electromyography (Noda et al., 2013; Pham et al., 2017) or similar sensing devices (Muratore et al., 2019), or humanoid robots (Lin et al., 2019; Zhou et al., 2019). Additionally, a robot might be able to enter environments that are unavailable to a user, allowing the user to teleoperate the robot in these environments, and effectively extend their reachable environment (Horiguchi et al., 2000). These strategies all effectively alter the user's mental state by decreasing the burden of user communication.

Finally, another strategy for robots to assist a user is by transforming the robot's own physical configuration into one that is more amenable to task completion. This approach is useful in collaborative scenarios where the robot and user may collide. To avoid this problem, robots can decrease their operating velocity when working in close proximity to users (Araiza-Illan and Clemente, 2018; Rosenstrauch et al., 2018; Svarny et al., 2019) or take paths or actions specifically designed to reduce the likelihood of a collision (De Luca and Flacco, 2012; Hayne et al., 2016; Liu et al., 2018; Nguyen et al., 2018). Similar to shared control, these strategies to assist the user decrease the user's cognitive burden of planning in the task. By taking responsibility for collisions, a robot can effectively alter its own actions so that the user can be less concerned with monitoring and modelling a robot's behavior and can concentrate on completing their portion of the task.

5.4 Implications

Choosing which action space the robot should act in is a crucial decision for robot designers. To aid users in room cleaning, for example, researchers have developed robots that alter the environment by directly picking up misplaced objects, while others have developed augmented reality solutions that provide assistance in the user's mental space by showing them routes that, if followed, would lead to the shortest time spent cleaning. Realizing that a given task can be solved by acting in any part of the state allows researchers to develop novel solutions to problems that have historically been restricted to robots that act in a single state.

This realization, however, means that determining the robot's action space is not simply determined by the task that the robot is being built to solve. Instead, a roboticist must carefully consider the capabilities of the users for whom they are designing the robot. The choice of how the robot acts must be tuned to the needs of the user, and it has broader implications on the user's sense of agency and trust in the system. This separation of robot action spaces enables designers to compare robots from different domains that have similar action spaces and develop better assistive solutions.

6 TIME

The third key design axis we present concerns how assistive robots coordinate the timing of actions with the targets of their assistance. Consider an assisted eating scenario. A robot might only offer food when given an explicit trigger by the user, or it can monitor the user's behavior to decide when to initiate the action itself. We categorize the timing of assistive actions as reactive, proactive, or simultaneous. Reactive robots act only when given explicit commands. Proactive robots use predictive models or other approaches to understand the world to initiate their actions without an explicit command. Robots acting simultaneously occur in collaborative settings, during which the robot continuously monitors the user for both explicit and implicit information to direct its actions. Choosing how to time the robot's behavior can change the difficulty of the task and how users react to the robot's assistance (**Figure 3**).

6.1 Reactive

Reactive assistance occurs when the assistive action is triggered by an explicit command. Consider a teleoperated robotic arm developed for assistive eating (Javdani et al., 2015; Aronson et al., 2018; Newman et al., 2018). In these studies, a user uses a two-degree of freedom joystick to control a seven-degree of freedom robot arm and pick up a morsel of food from a plate. Direct control of this robot entails only moving the robot's end-effector while the user is engaging the joystick. The user might also give commands at a higher level of abstraction, perhaps by pressing one button to request food and another for water.

Reactive robots can also respond to more task-specific, contextual triggers. In Canal et al. (2018), an assistive robot helps a user to put on their shoes. This interaction is modeled as a complicated handover problem, where the user must have their foot properly positioned and apply enough resistance that the shoe remains on the foot. In this work, the robot responds to a gesture performed by the user through their foot. When they move their

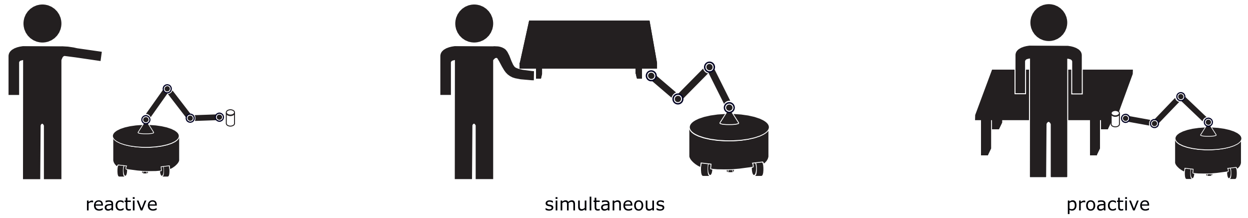


FIGURE 3 | A key axis in assistive robotic systems concerns what type of cue leads to the robot taking actions. Robots can be reactive and respond to explicit input only, be proactive and interpret the general task state to choose to act on their own, or collaborate closely with the user by acting simultaneously with them.

foot in the specified way, the robot knows that it is an acceptable time to place the shoe on their foot.

In general, reactive systems give the user more control over the robot and therefore agency in the overall interaction. Additionally, the robot does not generally need sophisticated models of the task, since it can rely on explicit input from the user. This simplicity means that the robot tends to be less sensitive to the particular task or domain, as it relies on the user to adapt the task to the robot's capabilities. However, this additional control requires the robot's user to spend more time and effort on controlling the robot, which can distract from other tasks. Controlling a robot at this level may also require significant training, as the robot's capabilities may not clearly match the requirements of the task. The control burden grows as the user must explicitly command the robot to begin an interaction (Baraglia et al., 2016), and requiring additional control complexity, such as adding modal control to teleoperation, can be cognitively taxing and slow down progress in the task (Herlant et al., 2016). Furthermore, requiring the user to explicitly cue the robot to act reduces collaborative fluency, which is undesirable as collaborative fluency is a positive attribute that has shown to increase the user's perceived quality of the interaction (Hoffman et al., 2014) and decrease the time spent during interactions (Huang and Mutlu, 2016).

6.2 Proactive

Proactive assistance occurs when the robot predicts that an action would fulfill the user's goals and takes that action without explicit prompting. For example, in assisted eating, the robot may anticipate a user's thirst after eating and choose to reach for the glass of water before receiving explicit input. The robot relies on a model of the task and user behavior to estimate what the user would want next. Proactive assistance generally improves the smoothness of interactions, as the assistance target does not need to spend time training or cognitive load to provide explicit instructions to the robot. However, this type of assistance is dependent on the model used to cue its actions, so the added complexity may make the system less reliable.

Consider again the task of operating a high degree of freedom robot using a low degree of freedom input device. Instead of using explicit signals from the user, Herlant et al. (2016) designed a robot that can proactively switch modes. In a simulated navigation task, a user drives a robot whose movement is restricted to exclusively moving either vertically or horizontally through a two-dimensional maze. The robot uses a model of the environment to determine whether horizontal or vertical motion is optimal given the robot's current position. The robot can then switch the mode proactively, allowing the user to simply direct the robot to move, speeding up the overall interaction time and removing the cognitive burden seen in reactive mode-switching.

Another way a robot can assist proactively is by building a model of the user to infer the task goal before it has been expressed. For example, a robot can predict the next fruit that a customer wants to add to their smoothie (Huang and Mutlu, 2016). Before the user explicitly requests this ingredient, the robot can prepare to grab that ingredient, increasing the fluidity of the interaction.

One challenge of proactive assistance is that users can be uncomfortable or even endangered if the robot makes unexpected motion. To mitigate this concern, the robot can communicate its intentions to the user explicitly. This could be done by having the robot show the user its plan directly on the physical environment, for example highlighting the part of a car door it plans to work on (Andersen et al., 2016b), or by showing its intended travel path in a virtual reality headset (Shu et al., 2018).

Proactive assistance enables more robust and general applications than reactive assistance does. However, the added sophistication in assistance requires additional complexity in the robot's models and behavior, which is compounded by the need to act in varied environments to unexpected stimuli. In addition, a purely proactive system can be uncomfortable or dangerous if the user is not prepared for the robot's actions. To mitigate some of these concerns, assistance systems can design some parts of the interaction as reactive and others as proactive. For example, the serving robot in Huang and Mutlu (2016) proactively moves closer to its estimate of the user's most likely request, but it does not initiate the actual grasping process until it receives an explicit command.

6.3 Simultaneous

Simultaneous assistance exists between the previous two categories and includes shared control and collaborative robots. These systems generally function similarly to proactive assistance, but act at the same time as the user. These systems include shared autonomy systems (Javdani et al., 2015; Javdani et al., 2018; Losey et al., 2018), which fuse the user's direct command with an autonomously generated command and arbitrate between the two according to some schema. It also includes tasks like carrying a table together (Nikolaidis et al., 2016; DelPreto and Rus, 2019), in which both the user and the robot must act independently for progress to be made.

Simultaneous assistance occurs often in collaborative assembly tasks. The goal and structure of a joint assembly task is often pre-specified, making it easy to determine a user's goal. A robot in such a task can directly assist by, for example, lifting and holding heavy objects steady so that they can be worked on (Fischer et al., 2015; El

Makrini et al., 2017). A robot can also assist by orienting a part to optimize construction, for example by following the images found in an assembly manual (Akkaladevi et al., 2016; Wang et al., 2020).

Simultaneous assistance often benefits from sophisticated communication strategies. For example, DelPreto and Rus (2019) designed a robot to sense electromyographic signals from a user to jointly manipulate a heavy object. A robot could also communicate back with the user, for example by changing its stiffness during a co-manipulation task in order to alert the user they should not move an object into a specific location (Bo et al., 2016). Similarly, a robot could provide the user with cues as to the next step during a complicated assembly task such as by pointing at the next item of interest (Admoni et al., 2016), providing a negative emotive feedback when a user completes an incorrect assembly step (Reyes et al., 2016; Rahman, 2019a) or display other emotive capabilities to signal task progress (Mok, 2016; Terzioglu et al., 2020).

Simultaneous assistive systems generally require tight collaboration between the user and the robot. The closeness of the collaboration requires the system to have a more complicated strategy for understanding user commands, since it is unlikely that the user will give precise commands while also accomplishing their task. However, these models can be more flexible than pure proactive systems: the robot can gain immediate feedback from the user about whether or not its action is correct, so it can recover from some model failures more quickly.

6.4 Implications

Determining when a robot should act has implications on the quality of a robot interaction. Reactive systems use more explicit control which enables more user agency, but it also increases the burden to complete a task. Proactive systems require more sophisticated models and sensing onboard the robot, but they can improve collaborative fluency while decreasing user burden. Systems that act in anticipation of explicit user commands may even be able to influence future user behavior in unforeseen ways, leading to questions about who is in control of setting the task goal (Newman et al., 2020). Proactive robots also generally lead to more robot agency, which introduces complex challenges such as safety and trust.

Preferences among when a robot chooses to take action may differ among users even within the same task domain. While one user may prefer a robot that requires less training and complication to operate, another might prefer to have more direct control over the robot to determine its behavior more precisely. If the user is paired with the system they least prefer, the interaction may cease to be assistive. In addition, an assistive system need not be completely proactive, reactive or simultaneous: the system can choose different timing and cueing strategies based on the particular part of the task under consideration. Choosing exactly when a robot executes its actions requires careful thought about the nature of the task, the capability of the robot, and the desires of the user.

7 CONCLUSION

In this paper, we describe an overall perspective on robotic systems that emphasizes their assistive intentions. With this perspective, we present three key design axes that compare assistive robotics research across domains: the relationships they develop with people, their action space, and their action timing. We explore these axes through a review of recent assistive robotics research, showing how assistive robots from across domains face similar challenges and make comparable decisions along these axes.

Much of the research discussed in this paper is specific to its task domain due to how the field has been organized and the difficulty of building abstractions. In this work, we propose some abstractions, and we hope that they will enable designers of assistive robots to find systems in other domains that share their problems and to draw deeper connections with them.

For each axis, we discuss design tradeoffs resulting from particular approaches. From among these axes, several themes emerge. Choices in the robot's action space and timing can both affect a user's sense of agency. Similarly, both the robot's action space and relationship with the user impact the structure of the communication between the robot and the user, which alters the quality of the assistance. It is our hope that researchers will explore more themes that span these design axes and provide more structure to the development of assistive robots.

Finally, this work is intended to start a conversation about how to understand the specific challenges of assistive robotics within the general area of human-robot interaction. With this framework, we hope to encourage researchers to further explore the nature of assistance as a general concept and describe its inherent challenges. We do not claim that these axes are complete; rather, we present them as the beginning of a larger effort to develop general principles of assistive robotics.

AUTHOR CONTRIBUTIONS

BAN, RMA, HA, and KK contributed to the conception and refinement of the main ideas of the paper. BAN developed the method for gathering papers for review. BAN read, selected, and organized the papers into the three critical axes. BAN wrote the first draft of the paper. BAN and RMA wrote the second draft of the paper, significantly reorganizing the first draft. RMA and BAN contributed to creating figures. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was partially supported by the National Science Foundation grant NSF IIS-1943072 and the Tang Family Foundation Innovation Fund. This research was also supported in part by a gift from Uptake, via the Uptake Machine Learning Fund for Social Good at Carnegie Mellon.

ACKNOWLEDGMENTS

We thank Alex London for discussions about the definitions and ethical implications of robotic assistance. We also thank Jessica Benner, Sarah Young, and Melanie Gainey for helping with the process of paper collection.

REFERENCES

Acharya, U., Kunde, S., Hall, L., Duncan, B. A., and Bradley, J. M. (2018). "Inference of User Qualities in Shared Control," in 2018 IEEE International Conference on Robotics and Automation (ICRA) (Brisbane, QLD, Australia: IEEE Computer Soc), 588–595. doi:10.1109/icra.2018.8461193

Ackovska, N., Kirandziska, V., Tanevska, A., Bozinovska, L., and Bozinovski, A. (2017). "Robot - Assisted Therapy for Autistic Children," in SoutheastCon 2017 (Concord, NC, USA: IEEE). doi:10.1109/secon.2017.7925401

Admoni, H., Weng, T., Hayes, B., and Scassellati, B. (2016). "Robot Nonverbal Behavior Improves Task Performance in Difficult Collaborations," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Christchurch, New Zealand: IEEE Press), 51–58. doi:10.1109/hri.2016.7451733

Agrigoroaie, R., Ferland, F., and Tapus, A. (2016). "The Enrichme Project: Lessons Learnt from a First Interaction with the Elderly," in *Social Robotics, (ICSR 2016)*. Editors A. Agah, J. Cabibihan, A. Howard, M. Salichs, and H. He (Berlin: Springer-Verlag), Vol. 9979, 735–745. doi:10.1007/978-3-319-47437-3_72

Akkaladevi, S. C., Plasch, M., Pichler, A., and Rinner, B. (2016). "Human Robot Collaboration to Reach a Common Goal in an Assembly Process," in *Proceedings of the Eighth European Starting Ai Researcher Symposium (Stairs 2016). Vol. 284 of Frontiers in Artificial Intelligence and Applications*. Editors D. Pearce and H. Pinto (Amsterdam: IOS Press), 3–14.

Alimardani, M., Kemmeren, L., Okumura, K., and Hiraki, K. (2020). "Robot-assisted Mindfulness Practice: Analysis of Neurophysiological Responses and Affective State Change," in 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 683–689. doi:10.1109/ro-man47096.2020.9223428

Andersen, R. S., Bogh, S., Moeslund, T. B., and Madsen, O. (2016a). "Task Space Hri for Cooperative Mobile Robots in Fit-Out Operations Inside Ship Superstructures," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 880–887. doi:10.1109/roman.2016.7745223

Andersen, R. S., Madsen, O., Moeslund, T. B., and Amor, H. B. (2016b). "Projecting Robot Intentions into Human Environments," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 294–301. doi:10.1109/roman.2016.7745145

Araiza-Illan, D., and Clemente, A. d. S. B. (2018). "Dynamic Regions to Enhance Safety in Human-Robot Interactions," in 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA) (IEEE), 693698.

Aronson, R. M., Santini, T., Kuebler, T. C., Kasneci, E., Srinivasa, S., and Admoni, H. (2018). "Eye-hand Behavior in Human-Robot Shared Manipulation," in HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (Assoc Computing Machinery) (IEEE), 4–13. doi:10.1145/3171221.3171287

Baklouti, M., Monacelli, E., Guitteny, V., and Couvet, S. (2008). "Intelligent Assistive Exoskeleton with Vision Based Interface," in *Smart Homes and Health Telematics*. *Vol. 5120 of Lecture Notes in Computer Science*. Editors S. Helal, S. Mitra, J. Wong, C. Chang, and M. Mokhtari (Berlin: Springer-Verlag), 123–135.

Baraglia, J., Cakmak, M., Nagai, Y., Rao, R., and Asada, M. (2016). "Initiative in Robot Assistance During Collaborative Task Execution," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (Christchurch, New Zealand: IEEE Press), 67–74. doi:10.1109/hri.2016.7451735

Bartl, A., Bosch, S., Brandt, M., Dittrich, M., and Lugrin, B. (2016). "The Influence of a Social Robot's Persona on How it Is Perceived and Accepted by Elderly Users," in *Social Robotics, (ICSR 2016). Vol. 9979 of Lecture Notes in Artificial Intelligence*. Editors A. Agah, J. Cabibihan, A. Howard, M. Salichs, and H. He (Berlin: Springer-Verlag), 681–691. doi:10.1007/978-3-319-47437-3_67

Bernardo, B., Alves-Oliveira, P., Santos, M. G., Melo, F. S., and Paiva, A. (2016). "An Interactive Tangram Game for Children with Autism," in *Intelligent Virtual Agents, IVA 2016. Vol. 10011 of Lecture Notes in Artificial Intelligence*. Editors D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (Berlin: Springer-Verlag), 500–504. doi:10.1007/978-3-319-47665-0_63

Bharatharaj, J., Huang, L., Al-Jumaily, A., Elara, M. R., and Krageloh, C. (2017). Investigating the Effects of Robot-Assisted Therapy Among Children with Autism Spectrum Disorder Using Bio-Markers. *IOP Conf. Series-Materials Sci. Eng.* 234, 012017. doi:10.1088/1757-899x/234/1/012017

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., and Sass, O. (2016). "Machines as a Source of Consolation: Robot Responsiveness Increases Human Approach Behavior and Desire for Companionship," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (Christchurch, New Zealand: IEEE Press), 165–171. doi:10.1109/hri.2016.7451748

Bo, H., Mohan, D. M., Azhar, M., Sreekanth, K., and Campolo, D. (2016). "Human-robot Collaboration for Tooling Path Guidance," in 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BIOROB) (Singapore: IEEE), 1340–1345. doi:10.1109/biorob.2016.7523818

Bodden, C., Rakita, D., Mutlu, B., and Gleicher, M. (2016). "Evaluating Intent-Expressive Robot Arm Motion," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE) (IEEE RO-MAN), 658663. doi:10.1109/roman.2016.7745188

Bonani, M., Oliveira, R., Correia, F., Rodrigues, A., Guerreiro, T., and Paiva, A. (2018). "What My Eyes Can't See, a Robot Can Show Me: Exploring the Collaboration between Blind People and Robots," in ASSETS '18: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY: Association for Computing Machinery), 15–27.

Breazeal, C., Brooks, A., Chilongo, D., Gray, J., Hoffman, G., Kidd, C., et al. (2004). "Working Collaboratively with Humanoid Robots," in 2004 4th IEEE/RAS International Conference on Humanoid Robots, Vols 1 and 2, Proceedings (IEEE), 253–272.

Briggs, P., Scheutz, M., and Tickle-Degnen, L. (2015). "Are Robots Ready for Administering Health Status Surveys? First Results from an Hri Study with Subjects with Parkinson's Disease," in Proceedings of the 2015 ACM/IEEE International Conference on Human-robot Interaction (HRI'15) (Assoc Computing Machinery) (IEEE), 327–334.

Broehl, C., Nelles, J., Brandl, C., Mertens, A., and Schlick, C. M. (2016). "Tam Reloaded: A Technology Acceptance Model for Human-Robot Cooperation in Production Systems," in *HCI International 2016 - Posters' Extended Abstracts, PT I. Vol. 617 of Communications in Computer and Information Science*. Editor C. Stephanidis (New York City: Springer International Publishing AG), 97–103.

Brose, S. W., Weber, D. J., Salatin, B. A., Grindle, G. G., Wang, H., Vazquez, J. J., et al. (2010). The Role of Assistive Robotics in the Lives of Persons with Disability. *Am. J. Phys. Med. Rehabil.* 89, 509–521. doi:10.1097/phm.0b013e3181cf569b

Bui, H.-D., and Chong, N. Y. (2018). "An Integrated Approach to Human-Robot-Smart Environment Interaction Interface for Ambient Assisted Living," in 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO) (IEEE), 32–37. doi:10.1109/arso.2018.8625821

Cacace, J., Caccavale, R., Finzi, A., and Lippiello, V. (2019a). "Variable Admittance Control Based on Virtual Fixtures for Human-Robot Co-manipulation," in 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (IEEE), 1569–1574.

Cacace, J., Finzi, A., and Lippiello, V. (2019b). "Enhancing Shared Control via Contact Force Classification in Human-Robot Cooperative Task Execution," in *Human Friendly Robotics. Vol. 7 of Springer Proceedings in Advanced Robotics*. Editors F. Ficuciello, F. Ruggiero, and A. Finzi (New York City: Springer International Publishing AG), 167–179. doi:10.1007/978-3-319-89327-3_13

Cakmak, M., Srinivasa, S. S., Lee, M. K., Forlizzi, J., and Kiesler, S. (2011). "Human Preferences for Robot-Human Hand-Over Configurations," in 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE). doi:10.1109/iros.2011.6094735

Canal, G., Alenya, G., and Torras, C. (2016). "Personalization Framework for Adaptive Robotic Feeding Assistance," in *Social Robotics, (ICSR 2016). Vol. 9979 of Lecture Notes in Artificial Intelligence*. Editors A. Agah, J. Cabibihan, A. Howard, M. Salichs, and H. He (Berlin: Springer-Verlag), 22–31. doi:10.1007/978-3-319-47437-3_3

Canal, G., Pignat, E., Alenya, G., Calinon, S., and Torras, C. (2018). "Joining High-Level Symbolic Planning with Low-Level Motion Primitives in Adaptive Hri: Application to Dressing Assistance," in 2018 IEEE International Conference on Robotics and Automation (ICRA) (Brisbane, QLD, Australia: IEEE), 3273–3278. doi:10.1109/icra.2018.8460606

Carmichael, M. G., and Liu, D. (2013). "Admittance Control Scheme for Implementing Model-Based Assistance-As-Needed on a Robot," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 870–873. doi:10.1109/embc.2013.6609639

Cha, E., Fitter, N. T., Kim, Y., Fong, T., and Mataric, M. J. (2018). "Effects of Robot Sound on Auditory Localization in Human-Robot Collaboration," in HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (Assoc Computing Machinery) (IEEE), 434–442. doi:10.1145/3171221.3171285

Chang, W.-L., and Sabanovic, S. (2015). "Interaction Expands Function: Social Shaping of the Therapeutic Robot Paro in a Nursing Home," in Proceedings of Tth 2015 ACM/IEEE International Conference on Human-Robot Interaction (HRI'15) (Assoc Computing Machinery) (IEEE), 343–350.

Chen, K., Plaza-Leiva, V., Min, B.-C., Steinfeld, A., and Dias, M. B. (2016). "Navcue: Context Immersive Navigation Assistance for Blind Travelers," in Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI'16) (Assoc Computing Machinery) (IEEE), 559. doi:10.1109/hri.2016.7451855

Chen, T. L., and Kemp, C. C. (2010). "Lead Me by the Hand: Evaluation of a Direct Physical Interface for Nursing Assistant Robots," in Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010) (IEEE), 367–374. doi:10.1109/hri.2010.5453162

Choi, H., Lee, J., and Kong, K. (2018). "A Human-Robot Interface System for Walkon Suit : a Powered Exoskeleton for Complete Paraplegics," in IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society (IEEE), 5057–5061. doi:10.1109/iecon.2018.8592884

Christodoulou, P., May Reid, A. A., Pnevmatikos, D., Rioja del Rio, C., and Fachantidis, N. (2020). Students Participate and Evaluate the Design and Development of a Social Robot 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) IEEE, 739–744. doi:10.1109/ro-man47096.2020.9223490

Cremer, S., Doelling, K., Lundberg, C. L., McNair, M., Shin, J., and Popa, D. (2016). "Application Requirements for Robotic Nursing Assistants in Hospital Environments," in *Sensors For Next-generation Robotics III. Vol. 9859 of Proceedings of SPIE*. Editors D. Popa and M. Wijesundara (Bellingham, Wa: Spie-Int Soc Optical Engineering). doi:10.1117/12.2229241

Cserteg, T., Erdos, G., and Horvath, G. (2018). "Assisted Assembly Process by Gesture Controlled Robots," in *51st CIRP Conference on Manufacturing Systems. Vol. 72 of Procedia CIRP*. Editor L. Wang (Amsterdam: Elsevier Science Bv), 51–56. doi:10.1016/j.procir.2018.03.028

Davison, D. P., Wijnen, F. M., Charisi, V., van der Meij, J., Evers, V., and Reidsma, D. (2020). "Working with a Social Robot in School: A Long-Term Real-World Unsupervised Deployment," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 63–72.

De Luca, A., and Flacco, F. (2012). "Integrated Control for Phri: Collision Avoidance, Detection, Reaction and Collaboration," in 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics And Biomechatronics (BIOROB). Editors J. Desai, L. Jay, and L. Zollo (Rome, Italy: IEEE), 288295. doi:10.1109/biorob.2012.6290917

DelPreto, J., and Rus, D. (2019). "Sharing the Load: Human-Robot Team Lifting Using Muscle Activity," in 2019 International Conference on Robotics and Automation (ICRA). Editors A. Howard, K. Althoefer, F. Arai, F. Arrichiello, B. Caputo, J. Castellanos, et al. (IEEE), 7906–7912. doi:10.1109/icra.2019.8794414

Devin, S., and Alami, R. (2016). "An Implemented Theory of Mind to Improve Human-Robot Shared Plans Execution," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 319–326. doi:10.1109/hri.2016.7451768

Dino, F., Zandie, R., Ahdollahi, H., Schoeder, S., and Mahoor, M. H. (2019). "Delivering Cognitive Behavioral Therapy Using a Conversational Social Robot," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 2089–2095. doi:10.1109/iros40897.2019.8968576

Dometios, A. C., Papageorgiou, X. S., Arvanitakis, A., Tzafestas, C. S., and Maragos, P. (2017). "Real-Time End-Effector Motion Behavior Planning Approach Using On-Line Point-Cloud Data Towards a User Adaptive Assistive Bath Robot," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Editors A. Bicchi and A. Okamura (IEEE), 50315036. doi:10.1109/iros.2017.8206387

Doroodgar, B., Ficocelli, M., Mobedi, B., and Nejat, G. (2010). "The Search for Survivors: Cooperative Human-Robot Interaction in Search and Rescue Environments Using Semi-Autonomous Robots," in 2010 IEEE International Conference on Robotics and Automation (ICRA) (IEEE), 2858–2863. doi:10.1109/robot.2010.5509530

Dragan, A. D., Lee, K. C. T., and Srinivasa, S. S. (2013). "Legibility and Predictability of Robot Motion," in 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 301–308. doi:10.1109/HRI.2013.6483603

El Makrini, I., Merckaert, K., Lefeber, D., and Vanderborght, B. (2017). "Design of a Collaborative Architecture for Human-Robot Assembly Tasks," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Editors A. Bicchi and A. Okamura (IEEE), 1624–1629. doi:10.1109/iros.2017.8205971

Fan, J., Beuscher, L., Newhouse, P. A., Mion, L. C., and Sarkar, N. (2016). "A Robotic Coach Architecture for Multi-User Human-Robot Interaction (Ramu) with the Elderly and Cognitively Impaired," in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 445–450. doi:10.1109/roman.2016.7745157

Faria, M., Silva, R., Alves-Oliveira, P., Melo, F. S., and Paiva, A. (2017). ""me and You Together" Movement Impact in Multi-User Collaboration Tasks," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Editors A. Bicchi and A. Okamura (IEEE), 2793–2798. doi:10.1109/iros.2017.8206109

Farrell, L., Strawser, P., Hambuchen, K., Baker, W., and Badger, J. (2017). "Supervisory Control of a Humanoid Robot in Microgravity for Manipulation Tasks," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Editors A. Bicchi and A. Okamura (IEEE), 37973802. doi:10.1109/iros.2017.8206229

Ferreira, B. Q., Karipidou, K., Rosa, F., Petisca, S., Alves-Oliveira, P., and Paiva, A. (2016). "A Study on Trust in a Robotic Suitcase," *Social Robotics, (ICSR 2016) Vol. 9979 of Lecture Notes in Artificial Intelligence*. Editors A. Agah, J. Cabibihan, A. Howard, M. Salichs, and H. He (Berlin: Springer-Verlag), 179–189. doi:10.1007/978-3-319-47437-3_18

Fischer, K., Jensen, L. C., Kirstein, F., Stabinger, S., Erkent, O., Shukla, D., et al. (2015). "The Effects of Social Gaze in Human-Robot Collaborative Assembly," *Social Robotics (ICSR 2015) Vol. 9388 of Lecture Notes in Artificial Intelligence* Editors A. Tapus, E. Andre, J. Martin, F. Ferland, and M. Ammi (Berlin: Springer-Verlag), 204–213. doi:10.1007/978-3-319-25554-5_21

Fong, T., Thorpe, C., and Baur, C. (2003). "Collaboration, Dialogue, and Human-Robot Interaction," *Robotics Research Vol. 6 of Springer Tracts in Advanced Robotics* Editors R. Jarvis and A. Zelinsky (Berlin: Springer-Verlag), 255–266.

Fuglerud, K. S., and Solheim, I. (2018). "The Use of Social Robots for Supporting Language Training of Children," in *Transforming Our World Through Design, Diversity And Education. Vol. 256 of Studies in Health Technology and Informatics*. Editors G. Craddock, C. Doran, L. McNutt, and D. Rice (Amsterdam, Netherlands: IOS PRESS), 401–408.

Gamborino, E., and Fu, L.-C. (2018). "Interactive Reinforcement Learning Based Assistive Robot for the Emotional Support of Children," in 2018 18th International Conference on Control, Automation and Systems (ICCAS) (IEEE), 708–713.

Gao, X., Gong, R., Zhao, Y., Wang, S., Shu, T., and Zhu, S.-C. (2020). "Joint Mind Modeling for Explanation Generation in Complex Human-Robot Collaborative Tasks," in 2020 29th Editors R. Jarvis and A. Zelinsky (Berlin: Springer-Verlag), 255 -266. IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1119–1126. doi:10.1109/ro-man47096.2020.9223595

Gao, Y., Sibirtseva, E., Castellano, G., and Kragic, D. (2019). "Fast Adaptation with Meta-Reinforcement Learning for Trust Modelling in Human-Robot Interaction," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE), 305–312. doi:10.1109/iros40897.2019.8967924

Goeruer, O. C., Rosman, B., Sivrikaya, F., and Albayrak, S. (2018). "Social Cobots: Anticipatory Decision-Making for Collaborative Robots Incorporating Unexpected Human Behaviors," in HRI '18: Proceedings of the 2018 ACM/

IEEE International Conference on Human-Robot Interaction (Assoc Computing Machinery (IEEE), 398–406.

Goldau, F. F., Shastha, T. K., Kyrarini, M., and Graeser, A. (2019). "Autonomous Multi-Sensory Robotic Assistant for a Drinking Task," in 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR) (IEEE), 210–216. doi:10.1109/icorr.2019.8779521

Goodrich, M. A., Colton, M., Brinton, B., and Fujiki, M. (2011). "A Case for Low-Dose Robotics in Autism Therapy," in Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2011) (IEEE), 143–144. doi:10.1145/1957656.1957702

Grames, E. M., Stillman, A. N., Tingley, M. W., and Elphick, C. S. (2019). An Automated Approach to Identifying Search Terms for Systematic Reviews Using Keyword Co-Occurrence Networks. *Methods Ecol. Evol.* 10, 1645–1654. doi:10.1111/2041-210X.13268

Grigore, E. C., Eder, K., Pipe, A. G., Melhuish, C., and Leonards, U. (2013). "Joint Action Understanding Improves Robot-To-Human Object Handover," in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Editor N. Amato (IEEE), 4622–4629. doi:10.1109/iros.2013.6697021

Guneysu, A., and Arnrich, B. (2017). "Socially Assistive Child-Robot Interaction in Physical Exercise Coaching," in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 670–675. doi:10.1109/roman.2017.8172375

Hawkins, K. P., Bansal, S., Vo, N. N., and Bobick, A. F. (2014). "Anticipating Human Actions for Collaboration in the Presence of Task and Sensor Uncertainty," in 2014 IEEE International Conference on Robotics and Automation (ICRA) (IEEE), 2215–2222. doi:10.1109/icra.2014.6907165

Hawkins, K. P., Vo, N., Bansal, S., and Bobick, A. F. (2013). "Probabilistic Human Action Prediction and Wait-Sensitive Planning for Responsive Human-Robot Collaboration," in 2013 13th IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS) (IEEE), 499–506. doi:10.1109/humanoids.2013.7030020

Hayne, R., Luo, R., and Berenson, D. (2016). "Considering Avoidance and Consistency in Motion Planning for Human-Robot Manipulation in a Shared Workspace," in 2016 IEEE International Conference on Robotics and Automation (ICRA). Editors A. Okamura, A. Menciassi, A. Ude, D. Burschka, D. Lee, F. Arrichiello, et al. (IEEE), 3948–3954. doi:10.1109/icra.2016.7487584

Hemminghaus, J., and Kopp, S. (2017). "Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI'17) (IEEE), 332–340. doi:10.1145/2909824.3020217

Herlant, L. V., Holladay, R. M., and Srinivasa, S. S. (2016). "Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 35–42. doi:10.1109/hri.2016.7451731

Hietanen, A., Changizi, A., Lanz, M., Kamarainen, J., Ganguly, P., Pieters, R., et al. (2019). "Proof of Concept of a Projection-Based Safety System for Human-Robot Collaborative Engine Assembly," in 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE). doi:10.1109/ro-man46459.2019.8956446

Hoffman, G., Birnbaum, G. E., Vanunu, K., Sass, O., and Reis, H. T. (2014). "Robot Responsiveness to Human Disclosure Affects Social Impression and Appeal," in HRI'14: Proceedings of the 2014 ACM/IEEE IInternational Conference on Human-robot Interaction (IEEE), 1–8. doi:10.1145/2559636.2559660

Horiguchi, Y., Sawaragi, T., and Akashi, G. (2000). "Naturalistic Human-Robot Collaboration Based upon Mixed-Initiative Interactions in Teleoperating Environment," in SMC 2000 Conference Proceedings: 2000 IEEE International Conference on Systems, Man & Cybernetics, Vol. 1-5 (IEEE), 876–881.

Huang, C.-M., and Mutlu, B. (2016). "Anticipatory Robot Control for Efficient Human-Robot Collaboration," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 83–90. doi:10.1109/hri.2016.7451737

Iossifidis, I., and Schoner, G. (2004). "Autonomous Reaching and Obstacle Avoidance with the Anthropomorphic Arm of a Robotic Assistant Using the Attractor Dynamics Approach," in 2004 IEEE International Conference on Robotics and Automation, Vols 1-5, Proceedings (IEEE), 4295–4300. doi:10.1109/robot.2004.1302393

Iqbal, T., and Riek, L. D. (2019). "Human-Robot Teaming: Approaches from Joint Action and Dynamical Systems," in *Humanoid Robotics: A Reference* (Netherlands: Springer Netherlands), 2293–2312. doi:10.1007/978-94-007-6046-2_137

Itadera, S., Kobayashi, T., Nakanishi, J., Aoyama, T., and Hasegawa, Y. (2019). "Impedance Control Based Assistive Mobility Aid through Online Classification of User's State," in 2019 IEEE/SICE International Symposium On System Integration (SII) (IEEE), 243–248. doi:10.1109/sii.2019.8700458

Ivanova, E., Krause, A., Schaelicke, M., Schellhardt, F., Jankowski, N., Achner, J., et al. (2017). "Let's Do This Together: Bi-manu-interact, a Novel Device for Studying Human Haptic Interactive Behavior," in 2017 International Conference on Rehabilitation Robotics (ICORR). Editors F. Amirabdollahian, E. Burdet, and L. Masia (IEEE), 708–713. doi:10.1109/icorr.2017.8009331

Jarrassé, N., Charalambous, T., and Burdet, E. (2012). A Framework to Describe, Analyze and Generate Interactive Motor Behaviors. *PLoS One* 7, e49945. doi:10.1371/journal.pone.0049945

Jarrasse, N., Paik, J., Pasqui, V., and Morel, G. (2008). "How Can Human Motion Prediction Increase Transparency?," in 2008 IEEE International Conference on Robotics AND Automation, Vols 1-9 (IEEE), 2134–2139.

Javdani, S., Admoni, H., Pellegrinelli, S., Srinivasa, S. S., and Bagnell, J. A. (2018). Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming. *Int. J. Robotics Res.* 37, 717–742. doi:10.1177/0278364918776060

Javdani, S., Srinivasa, S. S., and Bagnell, J. A. (2015). Shared Autonomy via Hindsight Optimization. *Robot Sci. Syst.* 2015.

Jensen, L. C., Fischer, K., Suvei, S.-D., and Bodenhagen, L. (2017). "Timing of Multimodal Robot Behaviors During Human-Robot Collaboration," in 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 1061–1066. doi:10.1109/roman.2017.8172435

Kanero, J., Franko, I., Oranc, C., Ulusahin, O., Koskulu, S., Adiguzel, Z., et al. (2018). "Who Can Benefit from Robots? Effects of Individual Differences in Robot-Assisted Language Learning," in 2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EPIROB) (IEEE), 212–217.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social Robot Tutoring for Child Second Language Learning," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 231–238. doi:10.1109/hri.2016.7451757

Kim, M.-G., Oosterling, I., Lourens, T., Staal, W., Buitelaar, J., Glennon, J., et al. (2014). "Designing Robot-Assisted Pivotal Response Training in Game Activity for Children with Autism," in 2014 IEEE International Conference on Systems, MAN and Cybernetics (SMC) (IEEE), 1101–1106. doi:10.1109/smc.2014.6974061

Knepper, R. A., Mavrogiannis, C. I., Proft, J., and Liang, C. (2017). "Implicit Communication in a Joint Action," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI'17) (IEEE), 283–292. doi:10.1145/2909824.3020226

Kontogiorgos, D., Pereira, A., Sahindal, B., van Waveren, S., and Gustafson, J. (2020). "Behavioural Responses to Robot Conversational Failures," in Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 53–62. doi:10.1145/3319502.3374782

Koppula, H. S., Jain, A., and Saxena, A. (2016). "Anticipatory Planning for Human-Robot Teams," in *Experimental Robotics. Vol. 109 of Springer Tracts in Advanced Robotics*. Editors M. Hsieh, O. Khatib, and V. Kumar (New York City: Springer International Publishing AG), 453–470. doi:10.1007/978-3-319-23778-7_30

Koskinopoulou, M., Piperakis, S., and Frahanias, P. (2016). "Learning from Demonstration Facilitates Human-Robot Collaborative Task Execution," in Eleventh ACM/IEEE IInternational Conference on Human Robot Interaction (HRI'16) (Assoc Computing Machinery) (IEEE), 59–66. doi:10.1109/hri.2016.7451734

Koustoumpardis, P. N., Chatzilygeroudis, K. I., Synodinos, A. I., and Aspragathos, N. A. (2016). "Human Robot Collaboration for Folding Fabrics Based on Force/

rgb-D Feedback," in *Advances in Robot Design and Intelligent Control. Vol. 371 of Advances in Intelligent Systems and Computing*. Editor T. Borangiu (Berlin: Springer-Verlag), 235–243. doi:10.1007/978-3-319-21290-6_24

Kratz, S., and Ferriera, F. R. (2016). "Immersed Remotely: Evaluating the Use of Head Mounted Devices for Remote Collaboration in Robotic Telepresence," in 2016 25th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 638–645. doi:10.1109/roman.2016.7745185

Kwon, W. Y., and Suh, I. H. (2012). "A Temporal Bayesian Network with Application to Design of a Proactive Robotic Assistant," in 2012 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE), 3685–3690. doi:10.1109/icra.2012.6224673

Kyrkjebo, E., Laastad, M. J., and Stavdahl, O. (2018). "Feasibility of the Ur5 Industrial Robot for Robotic Rehabilitation of the Upper Limbs after Stroke," in 2018 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS). Editors A. Maciejewski, A. Okamura, A. Bicchi, C. Stachniss, D. Song, D. Lee, et al. (IEEE), 6124–6129. doi:10.1109/iros.2018.8594413

Lambrecht, J., and Nimpsch, S. (2019). "Human Prediction for the Natural Instruction of Handovers in Human Robot Collaboration," in 2019 28th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE). doi:10.1109/ro-man46459.2019.8956379

Law, E., Cai, V., Liu, Q. F., Sasy, S., Goh, J., Blidaru, A., et al. (2017). "A Wizard-Of-Oz Study of Curiosity in Human-Robot Interaction," in 2017 26th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 607–614. doi:10.1109/roman.2017.8172365

Lee, H., and Hogan, N. (2016). "Essential Considerations for Design and Control of Human-Interactive Robots," in 2016 IEEE Internationl Conference on Robotics and Automation (ICRA). Editors A. Okamura, A. Menciassi, A. Ude, D. Burschka, D. Lee, F. Arrichiello, et al. (IEEE), 3069–3074. doi:10.1109/icra.2016.7487472

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2015). "Emotional Storytelling in the Classroom: Individual versus Group Interaction Between Children and Robots," in Proceedings of The 2015 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI'15) (Assoc Computing Machinery) (IEEE), 75–82.

Li, Y., Tee, K. P., Chan, W. L., Yan, R., Chua, Y., and Limbu, D. K. (2015). "Role Adaptation of Human and Robot in Collaborative Tasks," in 2015 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE Computer Soc) (IEEE), 5602–5607. doi:10.1109/icra.2015.7139983

Li, Z., and Hollis, R. (2019). "Toward a Ballbot for Physically Leading People: A Human-Centered Approach," in 2019 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS) (IEEE), 4827–4833. doi:10.1109/iros40897.2019.8968546

Lim, D., Kim, W., Lee, H., Kim, H., Shin, K., Park, T., et al. (2015). "Development of a Lower Extremity Exoskeleton Robot with a Quasi-Anthropomorphic Design Approach for Load Carriage," in 2015 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS) (IEEE), 5345–5350. doi:10.1109/iros.2015.7354132

Lin, T.-C., Krishnan, A. U., and Li, Z. (2019). "Physical Fatigue Analysis of Assistive Robot Teleoperation via Whole-Body Motion Mapping," in 2019 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS) (IEEE), 2240–2245. doi:10.1109/iros40897.2019.8968544

Liu, H., Wang, Y., Ji, W., and Wang, L. (2018). "A Context-Aware Safety System for Human-Robot Collaboration," in *28th Internationl Conference on Flexible Automation and Intelligent Manufacturing (FAIM2018): Global Integration Of Intelligent Manufacturing and Smart Industry for Good of Humanity. Vol. 17 of Procedia Manufacturing*. Editors D. Sormaz, G. Suer, and F. Chen (Elsevier Science Bv), 238–245. doi:10.1016/j.promfg.2018.10.042

Losey, D. P., McDonald, C. G., Battaglia, E., and O'Malley, M. K. (2018). A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction. *Appl. Mech. Rev.* 70. doi:10.1115/1.4039145

Louie, W.-Y. G., Vaquero, T., Nejat, G., and Beck, J. C. (2014). "An Autonomous Assistive Robot for Planning, Scheduling and Facilitating Multi-User Activities," in 2014 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE), 5292–5298. doi:10.1109/icra.2014.6907637

Lubold, N., Walker, E., and Pon-Barry, H. (2016). "Effects of Voice-Adaptation and Social Dialogue on Perceptions of a Robotic Learning Companion," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 255–262. doi:10.1109/hri.2016.7451760

Luo, J., Yang, C., Burdet, E., and Li, Y. (2020). "Adaptive Impedance Control with Trajectory Adaptation for Minimizing Interaction Force," in 2020 29th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1360–1365. doi:10.1109/ro-man47096.2020.9223572

Machino, T., Iwaki, S., Kawata, H., Yanagihara, Y., Nanjo, Y., and Shimokura, K.-i. (2006). "Remote-Collaboration System Using mobile Robot with Camera and Projector," in 2006 IEEE Internationl Conference on Robotics and Automation (ICRA), Vols 1-10 (IEEE), 4063+.

Maeda, G., Ewerton, M., Lioutikov, R., Ben Amor, H., Peters, J., and Neumann, G. (2014). "Learning Interaction for Collaborative Tasks with Probabilistic Movement Primitives," in 2014 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids) (IEEE), 527–534. doi:10.1109/humanoids.2014.7041413

Malik, N. A., Yussof, H., and Hanapiah, F. A. (2014). Development of Imitation Learning through Physical Therapy Using a Humanoid Robot. *Proced. Comput. Sci.* 42, 191–197. doi:10.1016/j.procs.2014.11.051

Malik, N. A., Yussof, H., and Hanapiah, F. A. (2017). Interactive Scenario Development of Robot-Assisted Therapy for Cerebral Palsy: A Face Validation Survey. *Proced. Comput. Sci.* 105, 322–327. doi:10.1016/j.procs.2017.01.229

Mariotti, E., Magrini, E., and De Luca, A. (2019). "Admittance Control for Human-Robot Interaction Using an Industrial Robot Equipped with a F/t Sensor," in 2019 Internationl Conference on Robotics and Automation (ICRA). Editors A. Howard, K. Althoefer, F. Arai, F. Arrichiello, B. Caputo, J. Castellanos, et al. (IEEE), 6130–6136. doi:10.1109/icra.2019.8793657

Martelaro, N., Nneji, V. C., Ju, W., and Hinds, P. (2016). "Tell Me More: Designing Hri to Encourage More Trust, Disclosure, and Companionship," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 181–188. doi:10.1109/hri.2016.7451864

Matarić, M. J., and Scassellati, B. (2016). *Socially Assistive Robotics*. Berlin, Germany: Springer Handbook of Robotics, 1973–1994.

Mataric, M., Tapus, A., Winstein, C., and Eriksson, J. (2009). "Socially Assistive Robotics for Stroke and Mild Tbi Rehabilitation," in *Advanced Technologies in Rehabilitation: Empowering Cognitive, Physical, Social and Communicative Skills Through Virtual Reality, Robots, Wearable Systems and Brain-computer Interfaces. Vol. 145 of Studies in Health Technology and Informatics*. Editors A. Gaggioli, E. Keshner, P. Weiss, and G. Riva (Amsterdam, Netherlands: IOS Press), 249–262.

Mertens, A., Reiser, U., Brenken, B., Luedtke, M., Haegele, M., Verl, A., et al. (2011). "Assistive Robots in Eldercare and Daily Living: Automation of Individual Services for Senior Citizens," in *Intelligent Robotics and Applications, PT I: ICIRA 2011. Vol. 7101 of Lecture Notes in Artificial Intelligence*. Editors S. Jeschke, H. Liu, and D. Schilberg (Berlin: Springer-Verlag), 542+. doi:10.1007/978-3-642-25486-4_54

Meyer, S., and Fricke, C. (2017). "Robotic Companions in Stroke Therapy: A User Study on the Efficacy of Assistive Robotics Among 30 Patients in Neurological Rehabilitation," in 2017 26th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 135–142. doi:10.1109/roman.2017.8172292

Milliez, G., Lallement, R., Fiore, M., and Alami, R. (2016). "Using Human Knowledge Awareness to Adapt Collaborative Plan Generation, Explanation and Monitoring," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 43–50. doi:10.1109/hri.2016.7451732

Miura, J., Kadekawa, S., Chikaarashi, K., and Sugiyama, J. (2016). "Human-Robot Collaborative Remote Object Search," in *Intelligent Autonomous Systems 13. Vol. 302 of Advances in Intelligent Systems and Computing*. Editors E. Menegatti, N. Michael, K. Berns, and H. Yamaguchi (Berlin: Springer-Verlag), 979–991. doi:10.1007/978-3-319-08338-4_71

Mizanoor Rahman, S. M. (2020). "Grasp Rehabilitation of Stroke Patients through Object Manipulation with an Intelligent Power Assist Robotic System," in *Proceedings of the Future Technologies Conference (FTC) 2019, Vol. 2. Vol. 1070 of Advances in Intelligent Systems and Computing*. Editors K. Arai, R. Bhatia, and S. Kapoor (New York City: Springer International Publishing Ag), 244–259. doi:10.1007/978-3-030-32523-7_16

Mok, B. (2016). "Effects of Proactivity and Expressivity on Collaboration with Interactive Robotic Drawers," in Eleventh ACM/IEEE Internationl Conference on Human Robot Interaction (HRI'16) (Assoc Computing Machinery) (IEEE), 633–634. doi:10.1109/hri.2016.7451892

Monaikul, N., Abbasi, B., Rysbek, Z., Di Eugenio, B., and Zefran, M. (2020). "Role Switching in Task-Oriented Multimodal Human-Robot Collaboration," in 2020 29th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1150–1156. doi:10.1109/ro-man47096.2020.9223461

Moon, H.-S., Kim, W., Han, S., and Seo, J. (2018). "Observation of Human Trajectory in Response to Haptic Feedback from Mobile Robot," in 2018 18th Internationl Conference on Control, Automation and Systems (ICCAS) (IEEE), 1530–1534.

Mucchiani, C., Sharma, S., Johnson, M., Sefcik, J., Vivio, N., Huang, J., et al. (2017). "Evaluating Older Adults' Interaction with a Mobile Assistive Robot," in 2017 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS). Editors A. Bicchi and A. Okamura (IEEE), 840–847. doi:10.1109/iros.2017.8202246

Mueller, S. L., Schroeder, S., Jeschke, S., and Richert, A. (2017). "Design of a Robotic Workmate," in *Digital Human Modeling: Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design*. *Vol. 10286 of Lecture Notes in Computer Science*. Editor V. Duffy (New York City: Springer International Publishing Ag), 447–456.

Muratore, L., Laurenzi, A., and Tsagarakis, N. G. (2019). "A Self-Modulated Impedance Multimodal Interaction Framework for Human-Robot Collaboration," in 2019 Internationl Conference on Robotics And Automation (ICRA). Editors A. Howard, K. Althoefer, F. Arai, F. Arrichiello, B. Caputo, J. Castellanos, et al. (IEEE), 4998–5004. doi:10.1109/icra.2019.8794168

Mussakhojayeva, S., Kalidolda, N., and Sandygulova, A. (2017). "Adaptive Strategies for Multi-Party Interactions with Robots in Public Spaces," in *Social Robotics, ICSR 2017*. *Vol. 10652 of Lecture Notes in Artificial Intelligence*. Editors A. Kheddar, E. Yoshida, S. Ge, K. Suzuki, J. Cabibihan, F. Eyssel, et al. (New York City: Springer International Publishing Ag), 749–758. doi:10.1007/978-3-319-70022-9_74

Muthusamy, R., Indave, J. M., Muthusamy, P. K., Hasan, E. F., Zweiri, Y., Kyrki, V., et al. (2019). "Investigation and Design of Robotic Assistance Control System for Cooperative Manipulation," in 2019 9th IEEE Annual Internationl Conference on Cyber Technology In Automation, Control, and Intelligent Systems (IEEE-Cyber 2019) (IEEE), 889–895. doi:10.1109/cyber46603.2019.9066573

Nabipour, M., and Moosavian, S. A. A. (2018). "Dynamics Modeling and Performance Analysis of Robowalk," in 2018 6th RSI Internationl Conference on Robotics and Mechatronics (ICROM 2018) (IEEE), 445–450. doi:10.1109/icrom.2018.8657593

Nakabayashi, K., Iwasaki, Y., Takahashi, S., and Iwata, H. (2018). "Experimental Evaluation of Cooperativeness and Collision Safety of a Wearable Robot Arm," in 2018 27th IEEE Internationl Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2018). Editors J. Cabibihan, F. Mastrogiovanni, A. Pandey, S. Rossi, and M. Staffa (IEEE), 1026–1031. doi:10.1109/roman.2018.8525620

Nemlekar, H., Dutia, D., and Li, Z. (2019). "Object Transfer Point Estimation for Fluent Human-Robot Handovers," in 2019 Internationl Conference on Robotics and Automation (ICRA). Editors A. Howard, K. Althoefer, F. Arai, F. Arrichiello, B. Caputo, J. Castellanos, et al. (IEEE), 2627–2633. doi:10.1109/icra.2019.8794008

Newman, B. A., Aronson, R. M., Srinivasa, S. S., Kitani, K., and Admoni, H. (2018). Harmonic: A Multimodal Dataset of Assistive Human-Robot Collaboration. *CoRR abs/1807.11154*.

Newman, B. A., Biswas, A., Ahuja, S., Girdhar, S., Kitani, K. K., and Admoni, H. (2020b). "Examining the Effects of Anticipatory Robot Assistance on Human Decision Making," in *Social Robotics*. Editors A. R. Wagner, D. Feil-Seifer, K. S. Haring, S. Rossi, T. Williams, H. He, et al. (New York City: Springer International Publishing), 590–603. doi:10.1007/978-3-030-62056-1_49

Newman, B., Carlberg, K., and Desai, R. (2020a). *Optimal Assistance for Object-Rearrangement Tasks in Augmented Reality*.

Nguyen, H., and Kemp, C. C. (2008). "Bio-Inspired Assistive Robotics: Service Dogs as a Model for Human-Robot Interaction and Mobile Manipulation," in 2008 2nd Ieee Ras & Embs Internationl Conference on Biomedical Robotics and Biomechatronics (BIOROB 2008), Vols 1 and 2 (IEEE), 542–549. doi:10.1109/biorob.2008.4762910

Nguyen, P. D. H., Bottarel, F., Pattacini, U., Hoffmann, M., Natale, L., and Metta, G. (2018). "Merging Physical and Social Interaction for Effective Human-Robot Collaboration," in 2018 IEEE-RAS 18th Internationl Conference on Humanoid Robots (Humanoids). Editor T. Asfour (IEEE), 710–717. doi:10.1109/humanoids.2018.8625030

Nie, G., Zheng, Z., Johnson, J., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., et al. (2018). "Predicting Response to Joint Attention Performance in Human-Human Interaction Based on Human-Robot Interaction for Young Children with Autism Spectrum Disorder," in 2018 27th IEEE Internationl Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2018). Editors J. Cabibihan, F. Mastrogiovanni, A. Pandey, S. Rossi, and M. Staffa (IEEE), 1069–1074. doi:10.1109/roman.2018.8525634

Nikolaidis, S., Kuznetsov, A., Hsu, D., and Srinivasa, S. (2016). "Formalizing Human-Robot Mutual Adaptation: A Bounded Memory Model," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 75–82. doi:10.1109/hri.2016.7451736

Noda, T., Furukawa, J.-i., Teramae, T., Hyon, S.-H., and Morimoto, J. (2013). "An Electromyogram Based Force Control Coordinated in Assistive Interaction," in 2013 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE), 2657–2662. doi:10.1109/icra.2013.6630942

Ono, H., Koike, K., Morita, T., and Yamaguchi, T. (2019). "Ontologies-Based Pupil Robot Interaction with Group Discussion," in *Knowledge-based and Intelligent Information & Engineering Systems (KES 2019). Vol. 159 of Procedia Computer Science*. Editors I. Rudas, C. Janos, C. Toro, J. Botzheim, R. Howlett, and L. Jain (Amsterdam: Elsevier), 2071–2080. doi:10.1016/j.procs.2019.09.380

Papageorgiou, X. S., Chalvatzaki, G., Dometios, A. C., and Tzafestas, C. S. (2019). "Human-Centered Service Robotic Systems for Assisted Living," in *Advances in Service and Industrial Robotics, Raad 2018. Vol. 67 of Mechanisms and Machine Science*. Editors N. Aspragathos, P. Koustoumpardis, and V. Moulianitis (New York City: Springer International Publishing Ag), 132–140. doi:10.1007/978-3-030-00232-9_14

Parlitz, C., Hägele, M., Klein, P., Seifert, J., and Dautenhahn, K. (2008). *Care-o-bot 3 — Rationale for Human-Robot Interaction Design*.

Peters, R., Broekens, J., Li, K., and Neerincx, M. A. (2019). "Robots Expressing Dominance: Effects of Behaviours and Modulation," in 2019 8th Internationl Conference on Affective Computing and Intelligent Interaction (ACII) (IEEE). doi:10.1109/acii.2019.8925500

Pham, T. X. N., Hayashi, K., Becker-Asano, C., Lacher, S., and Mizuuchi, I. (2017). "Evaluating the Usability and Users' Acceptance of a Kitchen Assistant Robot in Household Environment," in 2017 26th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 987–992. doi:10.1109/roman.2017.8172423

Polishuk, A., and Verner, I. (2018). "An Elementary Science Class with a Robot Teacher," in *Robotics in Education: Latest Results and Developments*. *Vol. 630 of Advances in Intelligent Systems and Computing*. Editors W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh, and D. Obdrzalek (New York City: Springer International Publishing Ag), 263–273. doi:10.1007/978-3-319-62875-2_24

Pripfl, J., Koertner, T., Batko-Klein, D., Hebesberger, D., Weninger, M., Gisinger, C., et al. (2016). "Results of a Real World Trial with a Mobile Social Service Robot for Older Adults," in Eleventh ACM/IEEE Internationl Conference on Human Robot Interaction (HRI'16) (Assoc Computing Machinery) (IEEE), 497–498. doi:10.1109/hri.2016.7451824

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Racca, M., Kyrki, V., and Cakmak, M. (2020). "Interactive Tuning of Robot Program Parameters via Expected Divergence Maximization," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 629–638. doi:10.1145/3319502.3374784

Rahman, S. M. M. (2019a). "Bioinspired Dynamic Affect-Based Motion Control of a Humanoid Robot to Collaborate with Human in Manufacturing," in 2019

12th Internationl Conference on Human System Interaction (HSI) (IEEE), 76–81. doi:10.1109/hsi47298.2019.8942609

Rahman, S. M. M. (2019b). "Human Features-Based Variable Admittance Control for Improving Hri and Performance in Power-Assisted Heavy Object Manipulation," in 2019 12th Internationl Conference on Human System Interaction (HSI) (IEEE), 87–92. doi:10.1109/hsi47298.2019.8942628

Ramachandran, A., Litoiu, A., and Scassellati, B. (2016). "Shaping Productive Help-Seeking Behavior During Robot-Child Tutoring Interactions," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 247–254. doi:10.1109/hri.2016.7451759

Reyes, M., Meza, I., and Pineda, L. A. (2016). "The Positive Effect of Negative Feedback in Hri Using a Facial Expression Robot," in *Cultural Robotics, CR 2015. Vol. 9549 of Lecture Notes in Computer Science*. Editors J. Koh, B. Dunstan, D. SilveraTawil, and M. Velonaki (New York City: Springer International Publishing Ag), 44–54. doi:10.1007/978-3-319-42945-8_4

Rhim, J., Cheung, A., Pham, D., Bae, S., Zhang, Z., Townsend, T., et al. (2019). "Investigating Positive Psychology Principles in Affective Robotics," in 2019 8th Internationl Conference on AFfective Computing and Intelligent Interaction (ACII) (IEEE). doi:10.1109/acii.2019.8925475

Riether, N., Hegel, F., Wrede, B., and Horstmann, G. (2012). "Social Facilitation with Social Robots?," in HRI'12: Proceedings of the Seventh Annual ACM/IEEE Internationl Conference on Human-Robot Interaction (Assoc Computing Machinery) (IEEE), 41–47.

Robinette, P., Li, W., Allen, R., Howard, A. M., and Wagner, A. R. (2016). "Overtrust of Robots in Emergency Evacuation Scenarios," in HRI '16: The Eleventh ACM/IEEE International Conference on Human Robot Interaction (IEEE Press), 101–108. doi:10.1109/hri.2016.7451740

Rosenberg-Kima, R., Koren, Y., Yachini, M., and Gordon, G. (2019). "Human-Robot-Collaboration (Hrc): Social Robots as Teaching Assistants for Training Activities in Small Groups," in HRI '19: 2019 14th ACM/IEEE Internationl Conference on Human-Robot Interaction (IEEE), 522–523. doi:10.1109/hri.2019.8673103

Rosenstrauch, M. J., Pannen, T. J., and Krueger, J. (2018). "Human Robot Collaboration - Using Kinect V2 for Iso/ts 15066 Speed and Separation Monitoring," in *7th Cirp Conference on Assembly Technologies and Systems (CATS 2018). Vol. 76 of Procedia CIRP*. Editors S. Wang and Z. Wu (Elsevier Science Bv), 183–186. doi:10.1016/j.procir.2018.01.026

Rossi, A., Dautenhahn, K., Koay, K. L., and Walters, M. L. (2020). "How Social Robots Influence People's Trust in Critical Situations," in 2020 29th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1020–1025. doi:10.1109/ro-man47096.2020.9223471

Sabanovic, S., Bennett, C. C., Chang, W.-L., and Huber, L. (2013). "Paro Robot Affects Diverse Interaction Modalities in Group Sensory Therapy for Older Adults with Dementia," in 2013 IEEE 13th Internationl Conference on Rehabilitation Robotics (ICORR) (IEEE). doi:10.1109/icorr.2013.6650427

Salichs, E., Fernández-Rodicio, E., Castillo, J. C., Castro-González, Á., Malfaz, M., and Salichs, M. Á. (2018). "A Social Robot Assisting in Cognitive Stimulation Therapy," in *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The Paams Collection. Vol. 10978 of Lecture Notes in Artificial Intelligence*. Editors Y. Demazeau, B. An, J. Bajo, and A. Fernandez Caballero (New York City: Springer International Publishing Ag), 344–347. doi:10.1007/978-3-319-94580-4_35

Sandygulova, A., Johal, W., Zhexenova, Z., Tleubayev, B., Zhanatkyzy, A., Turarova, A., et al. (2020). "Cowriting Kazakh: Learning a New Script with a Robot," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 113–120.

Savur, C., Kumar, S., and Sahin, F. (2019). "A Framework for Monitoring Human Physiological Response During Human Robot Collaborative Task," in 2019 IEEE Internationl Conference on Systems, Man and Cybernetics (SMC) (IEEE), 385–390. doi:10.1109/smc.2019.8914593

Schmidtler, J., and Bengler, K. (2016). "Size-Weight Illusion in Human-Robot Collaboration," in 2016 25th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 874–879. doi:10.1109/roman.2016.7745222

Schmidtler, J., Koerber, M., and Bengler, K. (2016). "A Trouble Shared is a Trouble Halved - Usability Measures for Human-Robot Collaboration," in 2016 IEEE Internationl Conference on Systems, Man, and Cybernetics (SMC) (IEEE), 217–222. doi:10.1109/smc.2016.7844244

Schneider, S., and Kummert, F. (2016). "Motivational Effects of Acknowledging Feedback from a Socially Assistive Robot," in *Social Robotics, (ICSR 2016). Vol. 9979 of Lecture Notes in Artificial Intelligence*. Editors A. Agah, J. Cabibihan, A. Howard, M. Salichs, and H. He (Berlin: Springer-Verlag), 870–879. doi:10.1007/978-3-319-47437-3_85

Sebo, S. S., Dong, L. L., Chang, N., and Scassellati, B. (2020). "Strategies for the Inclusion of Human Members within Human-Robot Teams," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 309–317. doi:10.1145/3319502.3374808

Sebo, S. S., Traeger, M., Jung, M., and Scassellati, B. (2018). "The Ripple Effects of Vulnerability: The Effects of a Robot's Vulnerable Behavior on Trust in Human-Robot Teams," in Hri '18: Proceedings of the 2018 ACM/IEEE Internationl Conference on Human-Robot Interaction (Assoc Computing Machinery) (IEEE), 178–186.

Shamaei, K., Che, Y., Murali, A., Sen, S., Patil, S., Goldberg, K., et al. (2015). "A Paced Shared-Control Teleoperated Architecture for Supervised Automation of Multilateral Surgical Tasks," in 2015 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS) (IEEE), 1434–1439. doi:10.1109/iros.2015.7353556

Shamsuddin, S., Zulkifli, W. Z., Hwee, L. T., and Yussof, H. (2017). "Animal Robot as Augmentative Strategy to Elevate Mood: A Preliminary Study for Post-Stroke Depression," in *Interactive Collaborative Robotics (ICR 2017). Vol. 10459 of Lecture Notes in Artificial Intelligence*. Editors A. Ronzhin, G. Rigoll, and R. Meshcheryakov (New York City: Springer International Publishing Ag), 209–218. doi:10.1007/978-3-319-66471-2_23

Shao, M., Alves, S. F. D. R., Ismail, O., Zhang, X., Nejat, G., and Benhabib, B. (2019). "You Are Doing Great! Only One Rep Left: An Affect-Aware Social Robot for Exercising," in 2019 IEEE Internationl Conference on Systems, Man and Cybernetics (SMC) (IEEE), 3811–3817. doi:10.1109/smc.2019.8914198

Sharifara, A., Ramesh Babu, A., Rajavenkatanarayanan, A., Collander, C., and Makedon, F. (2018). "A Robot-Based Cognitive Assessment Model Based on Visual Working Memory and Attention Level," *Universal Access in Human-Computer Interaction: Methods, Technologies, and Users, UAHCI 2018, PT IVol. 10907 of Lecture Notes in Computer Science.* Editors M. Antona and C. Stephanidis (New York City: Springer International Publishing Ag), 583–597. doi:10.1007/978-3-319-92049-8_43

Shayan, A. M., Sarmadi, A., Pirastehzad, A., Moradi, H., and Soleiman, P. (2016). "Roboparrot 2.0: A Multi-Purpose Social Robot," in 2016 4th RSI Internationl Conference on Robotics and Mechatronics (ICROM) (IEEE), 422–427.

Shibata, T., Mitsui, T., Wada, K., Touda, A., Kumasaka, T., Tagami, K., et al. (2001). "Mental Commit Robot and its Application to Therapy of Children," in 2001 IEEE/ASME Internationl Conference on Advanced Intelligent Mechatronics Proceedings, Vols I and II (IEEE), 1053–1058.

Shu, B., Sziebig, G., and Pieska, S. (2018). "Human-Robot Collaboration: Task Sharing through Virtual Reality," in IECOn 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society (IEEE), 6040–6044. doi:10.1109/iecon.2018.8591102

Sierra, S. D., Jimenez, M. F., Munera, M. C., Frizera-Neto, A., and Cifuentes, C. A. (2019). "Remote-Operated Multimodal Interface for Therapists During Walker-Assisted Gait Rehabilitation: A Preliminary Assessment," in HRI '19: 2019 14th ACM/IEEE Internationl Conference on Human-Robot Interaction (IEEE), 528–529. doi:10.1109/hri.2019.8673099

Sirithunge, H. P. C., Muthugala, M. A. V. J., Jayasekara, A. G. B. P., and Chandima, D. P. (2018). "A Wizard of Oz Study of Human Interest towards Robot Initiated Human-Robot Interaction," in 2018 27th IEEE Internationl Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2018). Editors J. Cabibihan, F. Mastrogiovanni, A. Pandey, S. Rossi, and M. Staffa (IEEE), 515–521. doi:10.1109/roman.2018.8525583

Sirkin, D., Mok, B., Yang, S., and Ju, W. (2015). "Mechanical Ottoman: How Robotic Furniture Offers and Withdraws Support," in 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 11–18.

Srinivasa, S. S., Berenson, D., Cakmak, M., Collet, A., Dogar, M. R., Dragan, A. D., et al. (2012). Herb 2.0: Lessons Learned from Developing a Mobile Manipulator for the Home. *Proc. IEEE* 100, 2410–2428. doi:10.1109/jproc.2012.2200561

Srinivasa, S. S., Ferguson, D., Helfrich, C. J., Berenson, D., Collet, A., Diankov, R., et al. (2010). Herb: a Home Exploring Robotic Butler. *Auton. Robot.* 28, 5–20. doi:10.1007/s10514-009-9160-9

Stoll, B., Reig, S., He, L., Kaplan, I., Jung, M. F., and Fussell, S. R. (2018). ""Wait, Can You Move the Robot?": Examining Telepresence Robot Use in Collaborative Teams," in HRI '18: Proceedings of the 2018 ACM/IEEE Internationl Conference on Human-Robot Interaction (Assoc Computing Machinery) (IEEE), 14–22.

Strohkorb, S., Fukuto, E., Warren, N., Taylor, C., Berry, B., and Seassellati, B. (2016). "Improving Human-Human Collaboration between Children with a Social Robot," in 2016 25th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN) (IEEE), 551–556. doi:10.1109/roman.2016.7745172

Su, H., Sandoval, J., Makhdoomi, M., Ferrigno, G., and De Momi, E. (2018). "Safety-Enhanced Human-Robot Interaction Control of Redundant Robot for Teleoperated Minimally Invasive Surgery," in 2018 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE Computer SOC) (IEEE), 6611–6616. doi:10.1109/icra.2018.8463148

Svarny, P., Tesar, M., Behrens, J. K., and Hoffmann, M. (2019). "Safe Physical Hri: Toward a Unified Treatment of Speed and Separation Monitoring Together with Power and Force Limiting," in 2019 IEEE/RSJ Internationl Conference on Intelligent Robots and Systems (IROS) (IEEE), 7580–7587. doi:10.1109/iros40897.2019.8968463

Tabrez, A., Agrawal, S., and Hayes, B. (2019). "Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning," in HRI '19: 2019 14th ACM/IEEE Internationl Conference on Human-Robot Interaction (IEEE), 249–257. doi:10.1109/hri.2019.8673104

Taheri, A. R., Alemi, M., Meghdari, A., PourEtemad, H. R., and Basiri, N. M. (2014). "Social Robots as Assistants for Autism Therapy in Iran: Research in Progress," in 2014 Second RSI/ISM Internationl Conference on Robotics and Mechatronics (ICROM) (IEEE), 760–766. doi:10.1109/icrom.2014.6990995

Tapus, A., Tapus, C., and Mataric, M. J. (2007). "Hands-off Therapist Robot Behavior Adaptation to User Personality for Post-Stroke Rehabilitation Therapy," in Proceedings of the 2007 IEEE Internationl Conference on Robotics and Automation, Vols 1-10 (IEEE), 1547+. doi:10.1109/robot.2007.363544

Terzioglu, Y., Mutlu, B., and Sahin, E. (2020). "Designing Social Cues for Collaborative Robots: The Role of Gaze and Breathing in Human-Robot Collaboration," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 343–357.

Till, J., Bryson, C. E., Chung, S., Orekhov, A., and Rucker, D. C. (2015). "Efficient Computation of Multiple Coupled Cosserat Rod Models for Real-Time Simulation and Control of Parallel Continuum Manipulators," in 2015 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE Computer Soc), IEEE International Conference on Robotics and Automation ICRA, 5067–5074. doi:10.1109/icra.2015.7139904

Torrey, C., Fussell, S. R., and Kiesler, S. (2013). "How a Robot Should Give Advice," in Proceedings of the 8th ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI 2013). Editors H. Kuzuoka, V. Evers, M. Imai, and J. Forlizzi (IEEE), 275–282. doi:10.1109/hri.2013.6483599

Uluer, P., Kose, H., Oz, B. K., Aydinalev, T. C., and Barkana, D. E. (2020). "Towards an Affective Robot Companion for Audiology Rehabilitation: How Does Pepper Feel Today?," in 2020 29th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 567–572.

Unhelkar, V. V., Li, S., and Shah, J. A. (2020). "Decision-making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 329–341. doi:10.1145/3319502.3374779

Unhelkar, V. V., Siu, H. C., and Shah, J. A. (2014). "Comparative Performance of Human and Mobile Robotic Assistants in Collaborative Fetch-and-Deliver Tasks," in HRI'14: Proceedings of the 2014 ACM/IEEE Internationl Conference on Human-Robot Interaction (IEEE), 82–89. doi:10.1145/2559636.2559655

van Minkelen, P., Gruson, C., van Hees, P., Willems, M., de Wit, J., Aarts, R., et al. (2020). "Using Self-Determination Theory in Social Robots to Increase Motivation in L2 Word Learning," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 369–377. doi:10.1145/3319502.3374828

Varrasi, S., Di Nuovo, S., Conti, D., and Di Nuovo, A. (2019). "Social Robots as Psychometric Tools for Cognitive Assessment: A Pilot Test," in *Human Friendly Robotics. Vol. 7 of Springer Proceedings in Advanced Robotics*. Editors F. Ficuciello, F. Ruggiero, and A. Finzi (New York City: Springer International Publishing Ag), 99–112. doi:10.1007/978-3-319-89327-3_8

Vatsal, V., and Hoffman, G. (2018). "Design and Analysis of a Wearable Robotic Forearm," in 2018 IEEE Internationl Conference on Robotics and Automation (ICRA) (IEEE Computer Soc) (IEEE), 5489–5496. doi:10.1109/icra.2018.8461212

Vishwanath, A., Singh, A., Chua, Y. H. V., Dauwels, J., and Magnenat-Thalmann, N. (2019). "Humanoid Co-workers: How is it like to Work with a Robot?," in 2019 28th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE).

Vu, D.-S., Allard, U. C., Gosselin, C., Routhier, F., Gosselin, B., and Campeau-Lecours, A. (2017). "Intuitive Adaptive Orientation Control of Assistive Robots for People Living with Upper Limb Disabilities," in 2017 Internationl Conference on Rehabilitation Robotics (ICORR). Editors F. Amirabdollahian, E. Burdet, and L. Masia (IEEE), 795–800. doi:10.1109/icorr.2017.8009345

Wandke, H. (2005). Assistance in Human–Machine Interaction: A Conceptual Framework and a Proposal for a Taxonomy. *Theoret. Iss. Ergonom. Sci.* 6, 129–155. doi:10.1080/1463922042000295669

Wang, Y., Ajaykumar, G., and Huang, C.-M. (2020). "See What I See: Enabling User-Centric Robotic Assistance Using First-Person Demonstrations," in Proceedings of the 2020 ACM/IEEE Internationl Conference on Human-Robot Interaction (HRI '20) (Assoc Computing Machinery) (IEEE), 639–648.

Westlund, J. K., Gordon, G., Spaulding, S., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Lessons from Teachers on Performing Hri Studies with Young Children in Schools," in Eleventh ACM/IEEE Internationl Conference on Human Robot Interaction (HRI'16) (Assoc Computing Machinery) (IEEE), 383–390. doi:10.1109/hri.2016.7451776

Wieser, I., Toprak, S., Grenzing, A., Hinz, T., Auddy, S., Karaoguz, E. C., et al. (2016). "Robotic Home Assistant with Memory Aid Functionality," in *KI 2016: Advances in Artificial Intelligence. Vol. 9904 of Lecture Notes in Artificial Intelligence*. Editors G. Friedrich, M. Helmert, and F. Wotawa (Berlin: Springer-Verlag), 102–115.

Winkle, K., and Bremner, P. (2017). "Investigating the Real World Impact of Emotion Portrayal through Robot Voice and Motion," in 2017 26th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 627–634. doi:10.1109/roman.2017.8172368

Wood, L. J., Zaraki, A., Walters, M. L., Novanda, O., Robins, B., and Dautenhahn, K. (2017). "The Iterative Development of the Humanoid Robot Kaspar: An Assistive Robot for Children with Autism," in *Social Robotics, Icsr 2017. Vol. 10652 of Lecture Notes in Artificial Intelligence*. Editors A. Kheddar, E. Yoshida, S. Ge, K. Suzuki, J. Cabibihan, F. Eyssel, et al. (New York City: Springer International Publishing Ag), 53–63. doi:10.1007/978-3-319-70022-9_6

Wu, M., He, Y., and Liu, S. (2020). "Shared Impedance Control Based on Reinforcement Learning in a Human-Robot Collaboration Task," in *Advances in Service and Industrial Robotics. Vol. 980 of Advances in Intelligent Systems and Computing*. Editors K. Berns and D. Gorges (New York City: Springer International Publishing Ag), 95–103. doi:10.1007/978-3-030-19648-6_12

Zafari, S., Schwaninger, I., Hirschmanner, M., Schmidbauer, C., Weiss, A., and Koeszegi, S. T. (2019). ""You are Doing So Great!" - the Effect of a Robot's Interaction Style on Self-Efficacy in Hri," in 2019 28th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE). doi:10.1109/ro-man46459.2019.8956437

Zhang, Z., Chen, Z., and Li, W. (2018). "Automating Robotic Furniture with a Collaborative Vision-Based Sensing Scheme," in 2018 27th IEEE Internationl Symposium on Robot and Human Interactive Communication (IEEE RO-MAN 2018). Editors J. Cabibihan, F. Mastrogiovanni, A. Pandey, S. Rossi, and M. Staffa (IEEE), 719–725. doi:10.1109/roman.2018.8525783

Zhao, F., Henrichs, C., and Mutlu, B. (2020). "Task Interdependence in Human-Robot Teaming," in 2020 29th IEEE Internationl Conference on Robot and Human Interactive Communication (RO-MAN) (IEEE), 1143–1149. doi:10.1109/ro-man47096.2020.9223555

Zhou, H., Yang, L., Lv, H., Yi, K., Yang, H., and Yang, G. (2019). "Development of a Synchronized Human-Robot-Virtuality Interaction System Using Cooperative Robot and Motion Capture Device," in 2019 IEEE/ASME Internationl Conference on Advanced Intelligent Mechatronics (AIM) (IEEE), 329–334. doi:10.1109/aim.2019.8868447

Zhu, H., Gabler, V., and Wollherr, D. (2017). "Legible Action Selection in Human-Robot Collaboration," in 2017 26th IEEE Internationl Symposium on Robot and Human Interactive Communication (RO-MAN). Editors A. Howard, K. Suzuki, and L. Zollo (IEEE), 354–359. doi:10.1109/roman.2017.8172326

Zignoli, A., Biral, F., Yokoyama, K., and Shimono, T. (2019). "Including a Musculoskeletal Model in the Control Loop of an Assistive Robot for the Design of Optimal Target Forces," in 45th Annual Conference of the IEEE Industrial Electronics Society (Iecon 2019) (IEEE), 5394–5400. doi:10.1109/iecon.2019.8927125

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Robotics and AI

REVIEW
published: 29 March 2022
doi: 10.3389/frobt.2022.704225

Still Not Solved: A Call for Renewed Focus on User-Centered Teleoperation Interfaces

Daniel J. Rea[1]* and Stela H. Seo[2]

[1]Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada, [2]Department of Social Informatics, Kyoto University, Kyoto, Japan

Teleoperation is one of the oldest applications of human-robot interaction, yet decades later, robots are still difficult to control in a variety of situations, especially when used by non-expert robot operators. That difficulty has relegated teleoperation to mostly expert-level use cases, though everyday jobs and lives could benefit from teleoperated robots by enabling people to get tasks done remotely. Research has made great progress by improving the capabilities of robots, and exploring a variety of interfaces to improve operator performance, but many non-expert applications of teleoperation are limited by the operator's ability to understand and control the robot effectively. We discuss the state of the art of user-centered research for teleoperation interfaces along with challenges teleoperation researchers face and discuss how an increased focus on human-centered teleoperation research can help push teleoperation into more everyday situations.

Keywords: teleoperation, interfaces, user-centered design, user-centered teleoperation, literature review, human-robot interaction

OPEN ACCESS

Edited by:
*Bilge Mutlu,
University of Wisconsin-Madison,
United States*

Reviewed by:
*Emmanuel Senft,
University of Wisconsin-Madison,
United States
Siddarth Jain,
Mitsubishi Electric Research
Laboratories (MERL), United States*

***Correspondence:**
*Daniel J. Rea
daniel.rea@unb.ca*

Specialty section:
*This article was submitted to
Human-Robot Interaction,
a section of the journal
Frontiers in Robotics and AI*

Received: *02 May 2021*
Accepted: *07 March 2022*
Published: *29 March 2022*

Citation:
Rea DJ and Seo SH (2022) Still Not Solved: A Call for Renewed Focus on User-Centered Teleoperation Interfaces. Front. Robot. AI 9:704225. doi: 10.3389/frobt.2022.704225

1 INTRODUCTION

Teleoperated robots, robots controlled at a distance, are already used in a number of situations where it is unsafe for humans to be physically present, such as search-and-rescue (Casper and Murphy, 2003; BBCClick, 2017; Peskoe-Yang, 2019) after natural disasters (Settimi et al., 2014; Norton et al., 2017; Tadokoro, 2019), scientific research such as moving underwater (Delmerico et al., 2019), working in space (Schilling et al., 1997), or making critical inspection and repairs (Buonocore, 2021; González et al., 2021; Koh et al., 2021). However, with a small number of exceptions, we see no remotely operated robots in use by the general public, whether it is at work or in their personal space. When we look to existing teleoperation situations–often robots in extreme situations–we find that teleoperation, even when performed by expert operators, is still an extremely difficult task, requiring multiple specialized operators for a single robot (Norton et al., 2017; Murphy and Tadokoro, 2019). Despite such human resources, it is still difficult to perform even basic collision avoidance, and this operation difficulty increases operator stress (Norton et al., 2017; Murphy and Tadokoro, 2019). If experts struggle to teleoperate robots remotely, then it is likely that the average person, even using simpler robots, would also struggle.

Teleoperation research has long noted that one of the bottlenecks to teleoperation performance is the operator's abilities, which can be limited by the technology itself, such as camera field of view (Endsley, 1988; Yanco et al., 2004; Chen et al., 2007; Lin et al., 2007; Niemeyer et al., 2016). In light of this, a long-term effective strategy for research is to create robots and interfaces that specifically reduce these limiting factors of the robot, such as adding additional camera views (Saakes et al., 2013;

Rakita et al., 2019a), leveraging multiple robots semi-autonomously (Rakita et al., 2019b; Coelho et al., 2021), or inventing new ways to control robots (Escolano et al., 2012; Choi et al., 2018). This research has resulted in numerous new techniques that can benefit numerous teleoperation scenarios. Despite the progress in these areas, teleoperation remains difficult.

In this paper, we review core challenges in teleoperation interface design and recent systematic surveys and find that teleoperation performance, especially for non-experts, is still hindered by the operator and the interface they use to control and monitor the robot. We found less user-centered work, instead focusing on improving and augmenting teleoperation technology to mitigate its weaknesses. User-centered work, in contrast, started from the abilities and needs of the operator and built interfaces with them in mind. Showcasing the potential of this approach, we highlight classic and recent examples for human-centered and non-expert teleoperation interface design [e.g., for manipulation (Herlant et al., 2016; Kent et al., 2017; Li et al., 2017; Rakita et al., 2017), shared autonomy (Jain et al., 2015; Aronson et al., 2018; Rakita et al., 2019b; Jeon et al., 2020; Nicolis, 2020), camera control (Rakita et al., 2018; Rakita et al., 2019a), and social and psychological interfaces (Rea et al., 2017a; Rea and Young, 2018; Rea and Young, 2019a; Rea et al., 2020; Seo et al., 2020; Valiton and Li, 2020)], we call for teleoperation and robot researchers broadly to use more additional advanced applications of user-centered practices by starting with user-driven solutions in addition to existing technically-driven approaches.

We further argue that engage with user-centered problems in teleoperation in a variety of applications, the field should focus more on simple everyday applications for non-expert users. We found that these seemingly simple tasks such as turning a door knob are still surprisingly difficult for modern teleoperation approaches, and we describe broad research directions for making user-centered interfaces as well as user-focused methods. Existing user-centered teleoperation research demonstrates that our call is a complimentary approach to traditional teleoperation research that has simply received less attention, but nevertheless has the potential for impact. This paper emphasizes and broadens other recent calls for increased focus on human factors (Murphy and Tadokoro, 2019), general usability (George et al., 2015), and information visualization (Szafir and Szafir, 2021). These directions can help bring teleoperated robots into daily life to improve productivity, accessibility, and more.

FIGURE 1 | Cockpit for remote drone teleoperation. Pilots need to process numerous sensors while handling mission tasks with multiple controls in stressful situations. (wikipedia.org/wiki/File:6th_Reconnaissance_Squadron_-_Operator.jpg, public domain).

2 CORE PROBLEMS IN TELEOPERATION INTERACTION

Using recent systematic surveys as a base [e.g., general telerobotics (Niemeyer et al., 2016), interfaces for teleoperated manipulators (Young and Peschel, 2020), or field robotics (Liu and Nejat, 2013; Delmerico et al., 2019; Murphy and Tadokoro, 2019)], we informally surveyed teleoperation interface research in recent years, as well as more influential work from the past 2 decades. We targeted the keywords of "teleoperation," "interaction," "interface," and "user-centered," in our journal and conference searches, and excluded work that was primarily engineering, algorithmic, or expert use-cases such as teleoperated space or surgery robots. We found most work focused on aiding two major user-centered problems in both expert and non-expert teleoperation: situation awareness and robot control. Together, these problems create a significant cognitive burden on the operator, making teleoperation difficult, regardless of the task being done with the robot. We briefly describe and discuss these major problems, and highlight some of the larger approaches we found. In particular, we found some research highlighting the user-centered nature of these problems, and successes for user-centered solutions.

2.1 Visualization of the Remote Data for Teleoperation Awareness

The term, situation awareness, emerged from aviation psychology to describe the pilot's understanding of tactical flight operations (Durso and Gronlund, 1999), but it is applicable broadly to any cognitive activity and information processing, including teleoperation (Durso and Gronlund, 1999; Yanco and Drury, 2004a; Endsley, 2015). Situation awareness, in teleoperation, is an operator's ability to perceive, understand, and reason about the environment around them and around their remote robot (Endsley, 1988; Rakita et al., 2017; Rakita et al., 2018; Nicolis, 2020), which requires the operator to process a large amount of robot sensor data from the remote environment in real-time (Goodrich et al., 2013).

2.1.1 Sensors for Situation Awareness

Research has been improving an operator's ability to build and maintain a high level of situation awareness for decades (Endsley, 1988; Endsley, 2016; Niemeyer et al., 2016). A general first-step approach is to add sensors to enable the robot to provide some

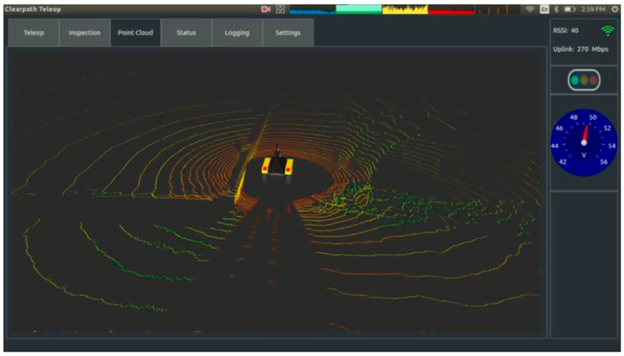


FIGURE 2 | A teleoperation interface from Clearpath Robotics. This tablet interface contains numerous sensor data streams which are accessed through mode switching via the top buttons in a tab-like interface. This interface reduces information load for the operator, but makes accessing all information quickly more difficult due to the need to switch tabs (source: Clearpath Robotics).

information for the operator to perceive the remote space and make better decisions [e.g., (Endsley, 1988; Endsley, 2016; Yanco and Drury, 2004a), see **Figures 1**, **2**]. Sensors may add perceptual abilities that people typically do not have, such as depth sensors that can see in the dark (Mast et al., 2015), or sonar that detects nearby objects (Nielsen et al., 2007). Additional sensors may instead simply add more advantageous data, such as egocentric cameras that provide a better detailed view, or exocentric cameras to provide a better sense of where the robot is in its environment (Saakes et al., 2013; Seo et al., 2017a). Entire other robots may be added to provide more abilities and viewpoints (Saakes et al., 2013; Rakita et al., 2019a; Coelho et al., 2021).

Adding more information via sensors, however, does not always help the operator's situation awareness, as people can only pay attention to a certain amount of information at once (Drury et al., 2003; Endsley, 2016; Norton et al., 2017). Instead of additional sensors, the user may be encouraged to perform robot actions to actively observe the environment [active perception (Bajcsy et al., 2018)]. This adds additional controls and perception needs to teleoperation, creating more cognitive load. The challenge of increasing cognitive load with robot capability is an on-going research topic in human-robot interaction (George et al., 2015; Delmerico et al., 2019; Murphy and Tadokoro, 2019). Research has therefore focused on reducing cognitive load for building and maintaining situation awareness to improve overall teleoperator performance.

2.1.2 Visualizations for Increasing Situational Awareness

To gain the benefits of additional sensors without increasing cognitive load, research is actively developing new techniques to visualize sensor data (Drury et al., 2003; Yanco and Drury, 2004b). A general trend is to process more raw data for the operation, and then create a display of those results, that is, easy for the operator to understand and reason about. For example, interfaces can highlight points of interest in ways that naturally draw the operator's interest (Rea et al., 2017b), leverage people's existing knowledge to summarize an off-screen's object's position and distance (Seo et al., 2017b), or combine multiple mediums like sound or haptics [multi-modal interfaces such as (Hacinecipoglu et al., 2013; Suarez Fernandez et al., 2016; Nicolis, 2020; Seo et al., 2020)]. Making robot sensor visualizations is not always obvious as sensors may detect qualities that are hard to visualize with a screen or speakers [such as haptic and force data (Reveleau et al., 2015; Nicolis, 2020)]. Interfaces like those described here—and many others (Nielsen et al., 2007; Yang et al., 2015; Seo et al., 2017a)—that fuse and interpret sensor data are essentially performing situation awareness processing for the operator, instead of having the operator analyze data or separate visualizations to come to conclusions themselves. How to produce such visualizations is difficult and an on-going topic in teleoperation, requiring more research in information visualization (Szafir and Szafir, 2021).

2.2 Robot Control Interfaces

In addition to creating and maintaining situation awareness from sensor data, the operator must make decisions quickly for their tasks and provide commands to a robot using the available controls. Understanding how to control a robot is also difficult for operators. This may be adjusting the robot's body pose, such as moving a multi-jointed robot arm, or to help drive a robot through an environment. Control itself consists of many problems, including situation awareness, simplify control complexity, choosing levels of autonomy for an action, or dealing with physical problems like latency. In general, the control scheme must be clear so that operators can understand and reason about how to command a robot to complete a task they may have—known as a gulf of execution that the operator must cross with the help of good interaction design (Norman et al., 1986). We found research typically focuses on one of two problems: how the input is provided by the operator, and what level of automation the robots behaviors use.

2.2.1 Level of Automation for Controls

A fundamental choice for robot control interfaces is how autonomous the actions will be. Historically, this has been almost no automation, with operators manually controlling each motor [e.g., early arms seen in (Singh et al., 2013) or remote-controlled vehicles]. However, in recent years, research has increasingly focused on introducing more semi-autonomous behaviors, enabling simple inputs to result in complex behaviors (Singh et al., 2013; Herlant et al., 2016; Kent et al., 2017; Li et al., 2017; Rakita et al., 2017; Rakita et al., 2019a; Nicolis, 2020; Young and Peschel, 2020). In addition, high levels of automation are vital to assist teleoperators in managing multiple robots while performing their tasks (Squire and Parasuraman, 2010; Wong and Seet, 2017). Automation enables the operator to think less about the robot's computers, sensors, and joints, and more as a tool that can accomplish tasks, reducing cognitive load.

2.2.2 Freeing Operators With Semi-Autonomous Controls

While there are clear benefits to semi-autonomous controls, there are tradeoffs and other problems introduced. For example, consider

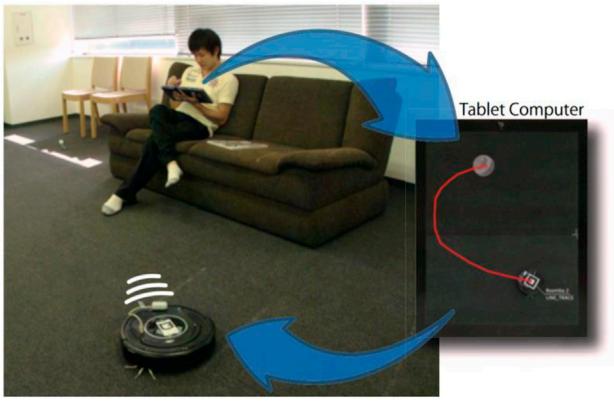


FIGURE 3 | A sketch-based control interface for a robot that overlays commands in an overhead view of the real world to aid control and understandability of the robot's future actions. From Sakamoto et al. (2009), with permission.

if the operator may define some level of goal (destination, pose, action, etc.), and then the robot autonomously proceeds partially or completely to that goal [e.g., (Quigley et al., 2004; Singh et al., 2013; Tsui et al., 2013)]. Once commands have been given, there is time while the robot proceeds which an operator can use to deal with other tasks [e.g., (Olsen and Wood, 2004; Glas et al., 2012)].

However, algorithms may be imperfect and the real world is dynamic, and so it may be necessary for the operator to provide more input during a task, such as help a kinematics simulator predict what position would be best to grip an option with a robotic arm (Leeper et al., 2012). Another potential drawback is that while attending other tasks, operators must maintain situation awareness of the teleoperation task, or reacquire it upon returning to the robot, potentially delaying task completion and adding workload to the operator (Donald et al., 2015). An operator may wish to edit or cancel an existing command in real-time, adding more complexity to the interaction. Because of the benefits for operator multitasking, and long-term planning of robot actions [e.g., (Liu et al., 2011; Sakamoto et al., 2009), see **Figure 3**], these problems for semi-autonomous teleoperation remain under active research, with research inventing new algorithms for autonomous behaviors and investigating their user acceptance [e.g., (Dragan and Srinivasa, 2013; Mehr et al., 2016; Javdani et al., 2018; Reddy et al., 2018; Brooks and Szafir, 2019)].

2.2.3 Robot and Mixed Initiative Controls

In addition to executing actions semi-autonomously, robots may also take actions by themselves (machine initiative) instead of waiting for operator commands [human initiative (Kortenkamp et al., 1997; Gombolay et al., 2017)]. This is particularly suited to be used in multiple robot teleoperation scenarios where one operator controls multiple robots [e.g., (Kolling et al., 2012; Glas et al., 2012)]. However, any autonomous action by the robot again threatens to break an operator's situation awareness and can also be confusing to the user if it is unclear the robot is taking initiative or a result of an operator command (mixed initiative systems). Such shared autonomy is a promising Frontier for improved usability and under active research [e.g., (Dragan and Srinivasa, 2013; Jain et al., 2015; Mehr et al., 2016; Aronson et al., 2018; Javdani et al., 2018; Reddy et al., 2018; Brooks and Szafir, 2019; Jeon et al., 2020), discussed later]. Recent research has shown the consistency and transparency in these robot-initiated actions is key to a better user experience (Small et al., 2018). Therefore, even with high levels of robot autonomy, we still need to consider the operator's user experience when creating teleoperation interfaces.

2.2.4 Input Strategies for Teleoperation

Even controlling a single robot is a challenging task that taxes an operator's cognitive resources (Steinfeld et al., 2006); seemingly basic tasks such as navigating a single, wheeled robot around a space are difficult enough that researchers have invented interfaces that aim to reduce the overhead required for a teleoperator in such a situation [e.g., (Barber et al., 2010; Young et al., 2011; Singh et al., 2013)].

2.2.5 Input Strategies for Non-Expert Teleoperators

Some strategies specifically targeting non-expert users (Kent et al., 2017; Li et al., 2017; Rakita et al., 2017; Rakita et al., 2019b; Jeon et al., 2020), such as employing well-known control metaphors [e.g., a dog leash for a robot companion (Young et al., 2011)], visualizing the results of a command (Singh et al., 2013), using intuitive controls such as sketching paths in an image of the environment (Sakamoto et al., 2009; Sugiura et al., 2010; Liu et al., 2011) (see **Figure 3**). These earlier works leverage user-centered design—the interface designs are rooted in familiar ways of acting and thinking (behavioral and cognitive psychology, human factors). We note that these controls simplify or reduce the degrees of freedom in the interface that the user has to explicitly think about (e.g., they simply move a pen instead of working with multi-button or multi-axis controls for 2D or 3D movement). With this approach of making interfaces more approachable, simpler, and familiar, the general public are more likely to find the controls usable than controls built for engineers or programmers (Singh et al., 2013; Delmerico et al., 2019).

2.2.6 Dynamic Control Granularity

Another strategy is to allow flexible levels of control. For example, an operator may need to define a precise path through an environment or grip an object at a certain angle; in these cases it is common to have complete control over robot movements with specialized interfaces designed for one robot's capabilities [e.g., (Sakamoto et al., 2009; Hashimoto et al., 2011; Glas et al., 2012)]. However, complex controls can make some actions, especially common actions, tiring to manually perform repeatedly. For these situations, one strategy is to combine those common but complex commands into single actions that are easy to invoke (Barber et al., 2010; Jain et al., 2015; Jeon et al., 2020)—once again simplifying the control space the user needs to think about. By understanding the tasks operators wish to complete with a robot, the interfaces can be made more manual or more automated to ease teleoperation.

While promising and demonstrably effective, we see little of these advances in modern teleoperation in our daily lives. It is possible that even further usability advances in feedback interfaces are needed to make teleoperation more accessible to the general public.

2.3 Robot Awareness of the Operator

An even less studied aspect of situation awareness that relates to controls is the robot's (and the teleoperation interface's) awareness of the operator. In other words, to properly execute commands and display appropriate feedback, teleoperation interfaces and robots should consider the environment and how commands serve the operator's goals (Endsley, 1988). A robot may also be aware of an operator's state, such as by user modelling—algorithmically predicting what a person is thinking, feeling, or will do (Basu et al., 2017; Jain and Argall, 2020; Jeon et al., 2020), or by directly monitoring the user [e.g., (Jia et al., 2014; Aronson et al., 2018)]. For example, control can be simplified by guessing operator intention to move to a space or pose and using automation to move towards that goal [e.g., (Gopinath et al., 2017; Aronson et al., 2018; Jain and Argall, 2020)], or the robot may understand a command but modify feedback to the operator to encourage better collision avoidance (Hacinecipoglu et al., 2013). Interfaces could even present information to create a specific state in the operator to affect future teleoperation (Rea and Young, 2019a; Rea et al., 2020).

Robots can also consider *how* an operator thinks a task should be completed, such as asking the operator to give input when an algorithm cannot decide between several courses of action [e.g., (Leeper et al., 2012)]. The robot should also be aware if the operator is distracted or busy with other tasks if input is needed, and use strategies such as taking intelligent automated actions until the operator is free [e.g., (Glas et al., 2012)], or the robot could use interfaces to draw operator attention quickly to important areas without distracting them too much (Rea et al., 2017b). Guessing operator intentions can be used to assess how much the operator has paid attention interface feedback, and could create an estimate of trust in operator commands.

Drawbacks and open challenges include how to integrate such machine initiative or shared autonomy actions in a way, that is, not disliked or confusing to operators (Wang and Lewis, 2007; Norton et al., 2017). Thus, by understanding the operator's state and goals, a robot can autonomously adapt commands to a dynamic world in an intelligent way.

The key idea is for the teleoperation system itself to consider not just the operator's commands, but their state—a user-aware interface—in order to help the operator's situation awareness and control accuracy. In other words, the operator is the center of the interface design. Because of this, we believe that even more user-centered designs and methodologies than are currently used are necessary for improving teleoperation.

3 USER-CENTERED PRACTICES AND TELEOPERATION

In our survey, we noted a trend for research to focus on solving core problems through additional robot capabilities and interface components, or addressing technical weaknesses (sensors, algorithms, etc.). We also found solutions driven by user needs and interface design in early and recent teleoperation research; these user-centered approaches show great promise, but we found fewer of these compared to technology-driven work. We discuss these and other recent works while asking–Why is teleoperation still so difficult?

3.1 Expert Interfaces for Usability Problems

Teleoperation has benefited from an increase in robot capabilities and robot-centered research, with improvements in reliability, robustness, sensing, traversal, and more (Delmerico et al., 2019; Murphy and Tadokoro, 2019). There was a general focus on expert users in highly technical, dangerous, and high-pressure situations, such as search and rescue use cases, bomb disposal and firefighting. In these cases, it is critical that the operator build an accurate mental picture of the situation quickly and control the robot successfully. Surveys noted that because the operator is so preoccupied with safe and careful operation, they often work in teams of multiple stakeholders (operators, engineers, decision-makers, etc.), that direct and advise the operator at an objective-level, rather than a robot level. This creates communication challenges between the types of operators, and researchers have noted they may each require a bespoke interface for optimal performance (Murphy and Tadokoro, 2019).

This leads to most modern teleoperation interfaces being expert interfaces—systems that assume extensive training and deep knowledge of a specialized system or application for good performance (Turchetti et al., 2012)[1]. These systems, due to their very specific user base, circumvent the need for common usability and learnability standards, often allowing or excusing increased information density, and complex, manual control. In this light, multiple operators may simply be one workaround to the usability difficulties of these systems. Both older (Endsley, 1988; Drury et al., 2003; Yanco et al., 2004; Chen and Thropp, 2007; Chen et al., 2011) and recent research (George et al., 2015; Murphy and Tadokoro, 2019; Szafir and Szafir, 2021), however, has identified that even these expert systems still have a need to incorporate and learn from basic HCI research (information density, learnability, layout, etc.) to further aid experts during teleoperation and decisions making, and going even further by developing and leveraging more advanced and specialized HCI areas like information visualization (Szafir and Szafir, 2021).

3.2 A Call for Additional Focus in User-Centered Teleoperation

While acknowledging that user-centered research has and continues to be an active subfield of teleoperation research [e.g., for manipulation (Jain et al., 2015; Herlant et al., 2016; Kent et al., 2017; Li et al., 2017; Rakita et al., 2017; Aronson et al., 2018; Rakita et al., 2019b), integrating automation (Jain et al., 2015; Rakita et al., 2019b; Jeon et al., 2020; Nicolis, 2020), or better sensor use (Rea et al., 2017b; Rakita et al., 2019a; Rea

[1]https://abcnews.go.com/Blotter/drone-stigma-means-skilled-pilots-controls-deadly-robots/story?id=23475968

and Young, 2019a; Rea et al., 2020; Seo et al., 2020)], based on our survey we call for more teleoperation researchers *to engage with teleoperation as a fundamentally user-driven domain* that should be approached with *user-centered processes*—a user evaluation alone, performed after development is only the beginning. Our goal should be that even everyday, non-expert people should be able to use complex robots with advanced manipulators, sensors, and movement abilities to improve everyday situations, and that *they should be included throughout the design and development process*. Even experts benefit from better user experience, usability, and learnability [e.g., in both general software (Dziubak et al., 2016) and robotics (Huang and Cakmak, 2017)], and user-focused improvements will lead to more accessible teleoperation in a variety of applications.

Recent research notes that teleoperation is fundamentally multitasking—doing a task with a robot, while figuring out how to accomplish that task safely and efficiently with the robot (Murphy and Tadokoro, 2019). Improving the fundamental basics of an interface (human factors, presentation, etc.) has long been known as an important factor in teleoperation (Yanco et al., 2004; Chen et al., 2007; Chen et al., 2011; George et al., 2015) but either sufficient improvement has yet to be made, or these basic HCI principles are insufficient on their own, perhaps due to the inherent complexities and difficulties of teleoperation (Norton et al., 2017; Norton et al., 2018; Murphy and Tadokoro, 2019; Szafir and Szafir, 2021). We propose that to conduct this user-focused teleoperation, more research should focus on general users in everyday applications for teleoperation, as it removes the ability to rely on expertise and technical ability.

To aid in this refocusing towards user-centered teleoperation, we highlight several applications of teleoperated robots that are not as extreme as the focus of many field robotics studies, note their unique problems that come from a more general user base, and motivate several future approaches for more research. We conclude that the shared nature of these problems with the core teleoperation problems described above suggests that teleoperation in general can progress by also investigating these simpler, more constrained applications, which in turn could provide new techniques and avenues for improvements in extreme robotics as well.

4 EVERYDAY APPLICATIONS OF TELEOPERATION—STILL NOT SOLVED

We have been arguing that teleoperation is fundamentally difficult from a usability perspective. To this end, we believe that researching interfaces tailored to everyday non-expert applications and users is important, as it provides similar research challenges in a simpler and more tractable testbed to progress fundamental usability issues in teleoperation. In fact, there have long been researchers that study everyday applications of teleoperation [e.g., (Niemeyer et al., 2016; Kent et al., 2017; Rakita et al., 2017)] where they encounter and study similar problems to more typical search-and-rescue robotics.

Telepresence can enter our daily lives in any situation a person needs to be in a physical space but cannot for any reason (health, distance, money, visas, etc.). Telepresence technologies has been used for children to attend school (Tsui et al., 2013), remote business meetings (Lee and Takayama, 2011), and more (Kristoffersson et al., 2013; Tsui et al., 2015). However, the core problems of teleoperation, difficult for experts, can be even more challenging for non-expert users. While well-developed commercial products exist, telepresence robots (Neustaedter et al., 2016; Rae and Neustaedter, 2017) are far from a solved problems, where challenges emerge from the surprising complexities of everyday situations and the interfaces needed to successfully navigate them.

Many industries have some level of routine inspection and maintenance needs that could be done with teleoperated robots. Some industries have well-defined and structured tasks and environments and tasks that could be leveraged as a simpler environment to develop better interfaces. Like more difficult applications of robots, these applications require robot navigation in constrained spaces, detailed inspection with multiple advanced sensors, logging, and reporting, and sometimes engaging in repair work using carefully controlled manipulators (Buonocore, 2021; González et al., 2021; Koh et al., 2021). These operators are often specialized in the industry the robot is used in, but not necessarily familiar with robotics or similar technology themselves. Thus, industrial teleoperation should benefit from increased user-centered research to aid these non-expert users while also acting as a simpler testbed for more complex types of teleoperation.

Teleoperated robots also have the potential to provide help in everyday situations, accessibility, and assist with ageing-in-place—assistive telerobotics (Goodrich et al., 2013; Jain et al., 2015; Tsui et al., 2015; Okamura and Tanaka, 2016; Jeon et al., 2020). For people who may have difficulties with mobility, strength, a comprised immune system (e.g., in a pandemic), or simply need assistance in a task, teleoperated robots could help by making tasks easier, reducing risk of injury or exposure to diseases. With improved interface design, teleoperated robots may improve feelings of efficacy, satisfaction, and independence of the home operators. One promising existing example of this technology is robotic wheelchairs with manipulators, which are not remote but still face typical core teleoperation challenges (Al-qaysi et al., 2018; Dragomir et al., 2021). These users may need extra help or have trouble using interfaces due to special needs, but designing for such users can improve customizability and accessibility for all users [e.g., (Jain et al., 2015; Jeon et al., 2020)].

Teleoperated robots may also increasingly become a form of everyday recreation [e.g., drone racing (Pfeiffer and Scaramuzza, 2021)]. The sport requires interfaces to support operators to drive safely but quicly while maintaining awareness in a highly dynamic environment. Drone racing is thus a useful real-world scenario to develop and test interfaces that help even everyday operators to leverage their skills to perceive, navigate, and interact with the environment in a highly dynamic situation. It further doubles as a safer test application for search and rescue robotics as fast-paced and difficult teleoperation situation, but with fewer serious consequences.

Looking back to core problems with teleoperation, we can see they all remain in everyday situations—achieving situation awareness with various sensors, and controls for navigation and manipulation. These tasks seem simple compared to the challenge of disaster response and search-and-rescue, but these "simple" everyday applications of teleoperation are still difficult from a usability perspective. We propose teleoperation requires research that focuses on the interaction design, and to develop more best practices and guidelines for user-centered interface design in teleoperation.

4.1 Considerations and Opportunities for Everyday User-Centered Teleoperation

The core user-centered teleoperation problems continue to be important to research from a systems perspective. However, with a user-centered approach, the research goals should shift to learnability, usability, and a better user experience, which can themselves increase operator performance, efficiency, and decision-making ability. We noted that everyday teleoperation applications provide good related and safer real-life situations to test and develop interfaces in. Here we discuss some main differences and opportunities in these applications: the operators, the robots, and the situations the robots are operated in.

4.1.1 The Operators

The operators are often experts in some field, but are not roboticists, engineers, or computer scientists. Thus, interfaces should not assume operators understand how to interpret a point cloud, how a differential drive works, or how kinematics affects how an arm can move; such topics will need to be avoided, hidden, or taught quickly but carefully in an interface. Operators may also not even be experts in their own specific robots, using them only a few times a year or less to do a maintenance task at their workplace. Everyday operators will also be more sensitive to user experience: since they are usually not required to use the robot, the robot must compete with the familiarity and ease of doing a task in a non-robotic way. They may even need additional accessibility considerations. The need to consider the lack of robot knowledge, familiarity, accessibility, and patience with poor user experience is not new in teleoperation research [e.g., (Drury et al., 2003; Chen et al., 2007; Sakamoto et al., 2009; Jain et al., 2015)], but we argue they need to become a core design consideration, as the experience and needs of operators differ heavily—additional use of user-centered research methods will be beneficial.

4.1.2 The Robots

The robots everyday people use may not be as advanced, robust, accurate, or powerful as those used by experts in extreme search-and-rescue situations, and this is potentially beneficial. For example, the commercial Double 3 telepresence robot does not have hundred-meter-reaching sensors, may not survive falling over or outdoor use, and does not move very quickly—we know how to build a more capable robot. However, these constraints came about from user-centered design: fewer robot features make creating a simple interface easier, it enables the robot to be built to suit a specific use case (e.g., indoor office or school use), and keep costs down for accessibility. In other words, a capable robot is not necessarily a better robot for a given user or application (Rea et al., 2017a). Leveraging these constraints and designing robots specifically with user needs in mind throughout the engineering process for more telepresence applications is an opportunity to improve robot accessibility and translate to better interfaces.

4.1.3 The Environment

The environments these robots are used in also may present specific challenges. For example, robots may be in dynamic or crowded spaces (a telepresence robot at a conference), or in an unknown and non-structured environment (a doctor tele-visiting a patient at their home). However, many environments are much more structured and regular than what search and rescue robots may be able to expect: factories have a well-known and predictable environment for inspection and maintenance robots, grocery stores have organized, easy-to-see-and-reach shelves of goods for a robot being used to pick up groceries, or public spaces like school or malls have a pre-defined environment. In addition, the tasks needed to be performed in these spaces may be completely or mostly defined in advance, such as a robot for routine maintenance. By understanding the user's needs in the task and environment, robots can be better designed to help in the most common cases.

4.1.4 Error Tolerance

Finally, while there are exceptions, many teleoperation applications are in situations that have some level of fault tolerance—delays will not result in lost lives, and minor collisions may only result in an apology and no harm. Thus, non-expert interfaces have an opportunity to help people learn how the mistakes they encountered came to be, and help avoid them in the future. This suggests that common performance metrics like number of errors may need to be rethought, and interfaces should explicitly be designed and tested expecting a certain level of operator error.

These considerations share the idea of simplification: less pre-supposed robot knowledge, simpler robots, simpler, safer, and more structured environments, and smaller costs for error. These simplifications may help make interface research more practical, while staying grounded in real-world problems, before extending to more difficult applications. While none of these problems are new or unknown [see active research such as (Herlant et al., 2016; Kent et al., 2017; Li et al., 2017; Rakita et al., 2017; Rakita et al., 2018; Rakita et al., 2019a; Rakita et al., 2019b; Jeon et al., 2020; Nicolis, 2020; Rea et al., 2020)], we call for additional attention to these problem areas as they are well suited for studying general user needs and perspectives.

5 CURRENT USER-CENTERED APPROACHES

We emphasize that the current systems and usability work being done in teleoperation is valuable—there are still hard robotics problems that would benefit usability if they were solved, and fundamental human factors work is still necessary as robotic platforms and new interaction methods (e.g., augmented or virtual reality) arise. In particular, we want to highlight how some existing work, some mentioned previously in **Section 2**, is user-focused and how those processes and techniques benefit teleoperation technologies.

5.1 Interfaces for Feedback

Research has developed many techniques over decades for displaying sensor data in an intuitive way, but the situation awareness problem is still unsolved (Norton et al., 2017; Norton et al., 2018; Delmerico et al., 2019; Murphy and Tadokoro, 2019). This is partly because new types of robot platforms and sensors have appeared, and it has become common for robots to have multiple sensors and cameras at once, increasing the operator's potential mental workload. If expert users struggle to interpret multiple sources of advanced sensor data, then there is a further need for simplified and easy to interpret data displays for non-expert operators.

Recent calls for bringing in advanced information visualization techniques acknowledge this, and is an important future approach to improving teleoperation (Szafir and Szafir, 2021). Some of these techniques even leverage new display technologies like augmented and virtual reality to explore new ways to present interfaces in a natural way (Nielsen et al., 2007; Labonte et al., 2010; Hedayati et al., 2018; González et al., 2021). We would encourage additional focus on visualizations that consider less experienced users, who may not understand or be familiar with the limitations of sensor or data types, who will likely be slower with any technique, and may be harder to train due to lack of foundational knowledge in engineering, or computer science.

Other recent surveys have noted that fundamental and basic human-computer interaction principles, such as limiting simultaneous information and considering interface layout, use of colors, and more are important (Niemeyer et al., 2016; Delmerico et al., 2019; Murphy and Tadokoro, 2019; Young and Peschel, 2020; Szafir and Szafir, 2021). Our survey agrees, and we encourage the development of additional user-centered teleoperation guidelines with additional and more in-depth user-driven solutions targeting non-expert applications.

5.2 Controls

Controls have also improved in terms of usability. Both aspects of the core problems discussed in **Section 2** have improved with user-focused methods: the physical hardware of the controller, and the way software can be used to add additional controls or abstract actions.

5.2.1 Simplifying With Abstraction and Automation

We observed a general trend for human-centered systems to keep controls simple by providing higher-level controls compared to traditional teleoperation systems (Sakamoto et al., 2009; Leeper et al., 2012; Ochiai et al., 2014; Kent et al., 2017), such as a point-and-click system navigation, with more complex controls being hidden with modes and menus. For more complex robots, the basic controls will often be modal, with the most common controls all accessible with video game or 3D haptic controllers, with perhaps advanced or more specific manual controls requiring something like a keyboard and mouse. However, there are many more advances for complex telemanipulation (Herlant et al., 2016; Kent et al., 2017; Li et al., 2017; Rakita et al., 2017; Rakita et al., 2018; Rakita et al., 2019a; Nicolis, 2020).

We also saw a trend for more abstract controls with automation, which can help non-expert operators with less experience. The key approach here is leveraging some level of autonomy to enable the operators to think at a task level (e.g., "grab this", "move there", "stop and inspect on the left"), rather than needing to reason about robot morphology, motors, or other low-level robot factors (Rea et al., 2020). This can relieve the workload of the operator [e.g., **Section 2.2.2** (Dragan and Srinivasa, 2013; Mehr et al., 2016; Javdani et al., 2018; Reddy et al., 2018; Brooks and Szafir, 2019)], though it may also create new user-centered problems related to initiative and transparency. This research is on-going and is necessary to both experts, and non-experts. In fact, being unable to rely on expertise may require even more clever displays of information and streamlined controls.

5.2.2 Simplifying With Modal Interfaces

We note some successes with modal inputs - the system's state changes how an input results in an action. For example, a joystick may normally move the robot forward and back, but the operator could enter a "manipulation mode" where the joystick instead moves a robot arm forward and back. Traditionally in human-computer interaction, modes are considered less usable for general software due to mistaking which mode the system is (Sarter and Woods, 1995) (mode error).

The common alternative for complex systems like telerobotics which often have many degrees of freedom inherently, however, is to just have a complex interface with many widgets, buttons, and controls (Yanco et al., 2015; Norton et al., 2017). The example modal control above is one method of enabling a smaller set of inputs to cover a broader range of robot actions. While this increases the possibility of mode error, well designed modal controls could simplify the controls space enough to make a net usability gain. Thus, more user centered work is required to gracefully enable high degree of freedom control to simple interfaces with potentially limited inputs.

As an example, we note that mode switching is commonly used in video games as a way to circumvent controllers with limited degrees of freedom to control a complex avatar. While video games are not necessarily simple to control, they are evidence that good user (or player)-centered designs can mitigate the drawbacks of modal designs and limit mode errors. We encourage this approach teleoperation designs, as teleoperation and video games have been shown to share many design similarities (Richer and Drury, 2006; Rea, 2020). Looking at the academic research of games usability, it further suggests that teleoperation may have its own set of usability guidelines that may differ from general software, encouraging further exploration of fundamental user-centered teleoperation research.

6 RESEARCH DIRECTIONS FOR USER-CENTERED TELEOPERATION

We are calling for a renewed user-centered focus to teleoperation interfaces, especially for everyday applications for non-expert users. We acknowledge that there has always been user-centered research in teleoperation, however our survey found limited engagement with this approach, focusing more often on technical solutions to user-based problems in expert applications. We view additional user-centered research as complimentary to existing systems-focused research in teleoperation and it will help operators take full advantage of the hardware and algorithms being developed today. In fact, many of these recommendations still require significant technical contributions to enable these user-centered approaches. We have highlighted state of the art successful techniques that are already demonstrating the power of this approach and point the way for future directions in a user-centered teleoperation.

We acknowledge the considerable overlap between the following high-level research directions, but we recommend future works focus on:

6.1 Help the User Do

Robot control is a large and general problem. However, there is already evidence that consistent, reliable controls that are intuitive and engaging to use while also accomplishing high-level actions can improve teleoperation. Future interfaces for robot control should aim for the following goals:

6.1.1 More Abstract Controls

A key trend we see in controls is abstraction–enabling operators to think more at a conceptual level than a robot or hardware level. Leveraging partial automation such as shared autonomy (Dragan and Srinivasa, 2013; Mehr et al., 2016; Javdani et al., 2018; Reddy et al., 2018; Brooks and Szafir, 2019) or other forms of automation can enable operators to think at the task level, rather than at the robot level. Manual modes should be placed behind menus or modes for more rare, dexterous tasks.

6.1.2 Better Experiences—Consistent, Reliable, and Transparent Controls

A user needs to be able to predict how a command will be performed by the robot (consistency). When a command is executing, succeeds, fails, or needs user input, the system should communicate this to the user (transparency). These, along with other guidelines, create good user experiences (Forlizzi and Battarbee, 2004; Hassenzahl and Tractinsky, 2006), and enable an operator to act with confidence, be engaged, satisfaction, and willingness to use again. We saw very little recent user experience-focused teleoperation work, but is known to be important to teleoperation (George et al., 2015), and interactive applications in general (Benyon, 2019).

6.1.3 Model the User to Better Interpret Commands

To be user-centered, the system should first understand the user. We suggest teleoperation systems further explore user monitoring (with sensors) and user modelling (predict how they are feeling and thinking) and adjust the interface and interpretation of commands accordingly. For example, the robot could detect a non-expert user is nervous, and ignore sudden erratic commands that are caused by nerves, or detect a tired operator and slow them down while reducing information displays (cognitive load) to help them think clearer.

6.2 Help the User Understand

Situation awareness is not simply about providing more information, it is about combining that information and visualizing it in a way that helps the user think and act effectively which is not straightforward in remote teleoperation. While this has been noted in other works (Seo et al., 2017a; Rea et al., 2017b; Seo et al., 2017b; Rea and Young, 2018; Rea and Young, 2019b; Szafir and Szafir, 2021), our own literature search corroborates this goal, and should be emphasized and be applied broadly to teleoperation. Mental workload is a fundamental metric for evaluating interface designs as workload has been strongly linked to teleoperation performance, and so research continues to target interfaces that reduce it, as well as improve other performance metrics (e.g., awareness) while adding minimal additional workload.

6.2.1 Leverage Human Psychology to Help People Process Information Naturally

People naturally process information in certain ways, for example, movement on screen can automatically draw a user's attention (Teng et al., 2013) and people automatically process information encoded in social interfaces (Feldmaier et al., 2017; Rea et al., 2020). This incurs a lower cognitive load than a multi-step rational analysis, for example, a multi-dimensional graph, which can be slow or need training non-expert operators may not have. Thus, we recommend visualizations that replicate how operators naturally process information (Rea et al., 2017b; Rea and Young, 2019a; Rea et al., 2020) (familiar representations such as faces, maps, etc.), which can be used to sidestep difficult engineering problems like computer vision, making the operator and robot as a sort of team (Mingyue Ma et al., 2018).

6.2.2 Create Understanding for Users and Avoid Displaying Raw Data

Modern teleoperation systems are often designed as expert-user interfaces, and so commonly display large amounts of information. However, for better teleoperation for all users, we recommend interfaces should add *knowledge* instead of information, by processing raw data and presenting it in a form, that, is more relevant and useful to users (Szafir and Szafir, 2021). Alternatively, the system should predict and only display relevant information to an operator, perhaps by leveraging known properties of the task, environments, or users. This could limit workload and increase non-expert operator ability, while still allowing deeper and expert interfaces to be present in menus, hidden for when needed.

6.2.3 Encourage Active Perception

People can build a better understanding by actively asking questions and exploring a space through action—active perception, or thinking through embodiment (Klemmer et al., 2006; Wainer et al., 2006; Bajcsy et al., 2018). An interface could encourage a user to move to a space, use a robot sensor or manipulator in a new way to understand the environment, and generally explore through interaction instead of simply thinking about the robot and sensor outputs. This method can be tailored to guide non-expert users especially, who may feel more confident moving and interacting in a space like they naturally do.

6.3 Help the User Learn

Even with user-centered interfaces, some training is inevitable Software learning has suggested general approaches to reduce the need and cost of this training, and turn using the robot into a learning experience for less experienced users:

6.3.1 Require Minimal Upfront Knowledge and Training

New designs should require minimal upfront specialized knowledge, and instead teach on-the-fly or leverage common designs that users can leverage from their everyday experiences (e.g., video games). If specialized interfaces are used, consider explaining them as they become relevant. For example, a telepresence system could detect the robot is looking at someone who is talking towards the robot, but that person's face is not visible. This could activate a tutorial to move the robot's camera to better see them and engage in conversation, teaching a new skill in an applied environment.

6.3.2 Enable Review of Interfaces and Mistakes

Another goal to improve non-expert teleoperation is to provide avenues for building skill, maintaining skill, reviewing the interface, and preventing future mistakes, as mistakes will happen and can be more excusable in everyday situations. Thus, we recommend researching interfaces that help users understand *when* an error occurred, *how* it occurred, and even how to *prevent* it in the future. For example, if a user does make an error (a collision, or has to stop and reposition the robot, etc.), the system could explain parts of the interface that were meant to inform the user of the nearby object, or even show a replay of recent operation, pointing out where the mistake could have been prevented.

We note our research directions are themselves user-centered. When building a user interface, researchers should focus on what they want to aid operators with, what the outcomes should be, and include users in the design process (not just in the evaluation stage). Teleoperation is not only about making improving robot capabilities, it is also about improving people's ability to complete tasks with robots.

Our goal in the medium-term is for comfortable single-robot non-expert operation. While the current multi-expert team standard in search and rescue teleoperation may maximize a robot's lifesaving potential, everyday non-expert operators have relaxed performance requirements and penalties for mistakes. This provides opportunity to better explore how to reduce information intelligently, help semi-automate common robot tasks, and improve interface learning and training. Teams of expert operators may always be the most effective in critical situations, but striving for comfortable single-operation by non-experts can make robots more appealing and applicable to a variety of applications.

7 CONCLUSION: WHY IS TELEOPERATION STILL SO HARD?

Teleoperation research has made great progress over the decades, improving robots, reducing latency, improving basic interfaces, and more. However, despite cheaper, more capable robots and many applications that could benefit from teleoperation, teleoperation remains at the edges of expert and extreme use cases. We argue that this is in part because teleoperation is a fundamentally difficult task for operators, and more user-centered methods should be applied to research in all areas of teleoperation design, especially in the interface. We surveyed teleoperation papers and found progress on the core teleoperation problems of control and situation awareness, and recent surveys and techniques that demonstrate the benefits of user-centered design for teleoperation. We called for a renewed focus in broad, user-centered research goals to improve teleoperation interfaces in everyday applications for non-experts, and to develop better interfaces that leverage how operators understand, think about, and use teleoperated robots. This leads us to recommend that end-users should be included throughout the teleoperation research process, not just as a user study at the end of a project, and that experiments should take advantage of such end-users' approachable everyday environments as experiment settings to test teleoperation technologies in the real world. The results of this research should complement the existing research approaches and benefit teleoperation as a whole.

AUTHOR CONTRIBUTIONS

DR was the lead researcher on both a conceptual and writing level, performing the majority of the work for both aspects. SS, was a contributing author during the brainstorming and conceptualizing, and contributed to the writing.

REFERENCES

Al-qaysi, Z. T., Zaidan, B. B., Zaidan, A. A., and Suzani, M. S. (2018). A Review of Disability EEG Based Wheelchair Control System: Coherent Taxonomy, Open Challenges and Recommendations. *Comput. Methods Programs Biomed.* 164, 221–237. doi:10.1016/j.cmpb.2018.06.012

Aronson, R. M., Santini, T., Kübler, T. C., Kasneci, E., Srinivasa, S., and Admoni, H. (2018). "Eye-Hand Behavior in Human-Robot Shared Manipulation," in *Human-Robot Interaction*, 4–13. doi:10.1145/3171221.3171287

Bajcsy, R., Aloimonos, Y., and Tsotsos, J. K. (2018). Revisiting Active Perception. *Auton. Robot* 42 (2), 177–196. doi:10.1007/s10514-017-9615-3

Barber, C. M., Shucksmith, R. J., MacDonald, B., and Wünsche, B. C. (2010). "Sketch-based Robot Programming," in *25th International Conference of Image and Vision Computing New Zealand.*

Basu, C., Yang, Q., Hungerman, D., Singhal, M., and Dragan, A. D. (2017). "Do You Want Your Autonomous Car to Drive like You?," in *Human-Robot Interaction*, 417–425.

BBCClick (2017). Fukushima Disaster: The Robots Going where No Human Can. *October 18*.

Benyon, D. (2019). *Designing User Experience: A Guide to HCI, UX and Interaction Design*. Pearson.

Brooks, C., and Szafir, D. (2019). "Balanced Information Gathering and Goal-Oriented Actions in Shared Autonomy," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 11-14 March 2019, 85–94. doi:10.1109/HRI.2019.8673192

Buonocore, Luca. R. (2021). *Overview of the Robotic Service at CERN for Accelerator Maintenance*.

Casper, J., and Murphy, R. R. (2003). Human-robot Interactions during the Robot-Assisted Urban Search and rescue Response at the World Trade Center. *IEEE Trans. Syst. Man. Cybern. B* 33 (3), 367–385. doi:10.1109/tsmcb.2003.811794

Chen, J. Y. C., Barnes, M. J., and Harper-Sciarini, M. (2011). Supervisory Control of Multiple Robots: Human-Performance Issues and User-Interface Design. *IEEE Trans. Syst. Man. Cybern. C* 41 (4), 435–454. doi:10.1109/tsmcc.2010.2056682

Chen, J. Y. C., Haas, E. C., and Barnes, M. J. (2007). Human Performance Issues and User Interface Design for Teleoperated Robots. *IEEE Trans. Syst. Man. Cybern. C* 37 (6), 1231–1245. doi:10.1109/tsmcc.2007.905819

Chen, J. Y. C., and Thropp, J. E. (2007). Review of Low Frame Rate Effects on Human Performance. *IEEE Trans. Syst. Man. Cybern. A.* 37 (6), 1063–1076. doi:10.1109/tsmca.2007.904779

Choi, P. J., Oskouian, R. J., and Tubbs, S. R. (2018). Telesurgery: Past, Present, and Future. *Cureus* 10 (5), 1–5. doi:10.7759/cureus.2716

Coelho, A., Sarkisov, Y., Wu, X., Mishra, H., Singh, H., Dietrich, A., et al. (2021). Whole-Body Teleoperation and Shared Control of Redundant Robots with Applications to Aerial Manipulation. *J. Intell. Robotic Syst. Theor. Appl.* 102, 1. doi:10.1007/s10846-021-01365-7

Delmerico, J., Mintchev, S., Giusti, A., Gromov, B., Melo, K., Horvat, T., et al. (2019). The Current State and Future Outlook of rescue Robotics. *J. Field Robotics* 36 (7), 1171–1191. doi:10.1002/rob.21887

Donald, F., Donald, C., and Thatcher, A. (2015). Work Exposure and Vigilance Decrements in Closed Circuit Television Surveillance. *Appl. Ergon.* 47 (January 2016), 220–228. doi:10.1016/j.apergo.2014.10.001

Dragan, A. D., and Srinivasa, S. S. (2013). A Policy-Blending Formalism for Shared Control. *Int. J. Robotics Res.* 32 (7), 790–805. doi:10.1177/0278364913490324

Dragomir, A., Pană, C. F., Cojocaru, D., and Manta, L. F. (2021). "Human-Machine Interface for Controlling a Light Robotic Arm by Persons with Special Needs," in *2021 22nd International Carpathian Control Conference (ICCC)*.

Drury, J. L. L., Scholtz, J., and Yanco, H. A. (2003). "Awareness in Human-Robot Interactions," in SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483), Washington, DC, USA, 8-8 Oct. 2003, 912–918.

Durso, F. T., and Gronlund, S. D. (1999). "Situation Awareness," in *Handbook of Applied Cognition*. Editor F. T. Durso (John Wiley & Sons, Inc.), 283–314.

Dziubak, V., Dubois, P., Bunt, A., and Terry, M. (2016). "Switter: Supporting Exploration of Software Learning Materials on Social Media," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 1209–1220.

Endsley, M. R. (2016). *Designing for Situation Awareness: An Approach to User-Centered Design*. CRC Press.

Endsley, M. R. (1988). "Design and Evaluation for Situation Awareness Enhancement," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Sage 32, 97–101.

Endsley, M. R. (2015). Situation Awareness: Operationally Necessary and Scientifically Grounded. *Cogn. Tech. Work* 17 (2), 163–167. doi:10.1007/s10111-015-0323-5

Escolano, C., Antelis, J. M., and Minguez, J. (2012). A Telepresence mobile Robot Controlled with a Noninvasive Brain-Computer Interface. *IEEE Trans. Syst. Man. Cybern. B* 42 (3), 793–804. doi:10.1109/tsmcb.2011.2177968

Feldmaier, J., Stimpfl, M., and Diepold, K. (2017). "Development of an Emotion-Competent SLAM Agent," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. New York, NY: ACM.

Forlizzi, J., and Battarbee, K. (2004). "Understanding Experience in Interactive Systems," in Proceedings of the 2004 conference on Designing interactive systems processes, practices, methods, and techniques - DIS '04, Cambridge MA USA, August 1 - 4, 2004, 261. doi:10.1145/1013115.1013152

George, A., Christou, G., Katsanos, C., Xenos, M., and Hadzilacos, T. (2015). Usability Guidelines for the Design of Robot Teleoperation: A Taxonomy. *IEEE Trans. Human-Machine Syst.* 45 (2), 256–262. doi:10.1109/THMS.2014.2371048

Glas, D. F., Kanda, T., Ishiguro, H., and Hagita, N. (2012). Teleoperation of Multiple Social Robots. *IEEE Trans. Syst. Man. Cybern. A.* 42 (3), 530–544. doi:10.1109/tsmca.2011.2164243

Gombolay, M., Bair, A., Huang, C., and Shah, J. (2017). Computational Design of Mixed-Initiative Human–Robot Teaming that Considers Human Factors: Situational Awareness, Workload, and Workflow Preferences. *Int. J. Robotics Res.* 36 (5–7), 597–617. doi:10.1177/0278364916688255

González, C., Ernesto Solanes, J., Muñoz, A., Gracia, L., Girbés-Juan, V., and Tornero, J. (2021). Advanced Teleoperation and Control System for Industrial Robots Based on Augmented Virtuality and Haptic Feedback. *J. Manufacturing Syst.* 59 (February), 283–298. doi:10.1016/j.jmsy.2021.02.013

Goodrich, M. A., Crandall, J. W., and Barakova, E. (2013). Teleoperation and beyond for Assistive Humanoid Robots. *Rev. Hum. Factors Ergon.* 9 (1), 175–226. doi:10.1177/1557234x13502463

Gopinath, D., Jain, S., and Argall, B. D. (2017). Human-in-the-Loop Optimization of Shared Autonomy in Assistive Robotics. *IEEE Robot. Autom. Lett.* 2 (1), 247–254. doi:10.1109/lra.2016.2593928

Hacinecipoglu, A., Konukseven, E. I., and Bugra Koku, A. (2013). "Evaluation of Haptic Feedback Cues on Vehicle Teleoperation Performance in an Obstacle Avoidance Scenario," in *2013 World Haptics Conference, WHC (2013)*. IEEE.

Hashimoto, S., Ishida, A., Inami, M., and Igarash, T. (2011). "TouchMe: An Augmented Reality Based Remote Robot Manipulation," in *International Conference on Artificial Reality and Telexistence*. Osaka, Japan.

Hassenzahl, M., and Tractinsky, N. (2006). User Experience - A Research Agenda. *Behav. Inf. Techn.* 25 (2), 91–97. doi:10.1080/01449290500330331

Hedayati, H., Walker, M., and Szafir, D. (2018). "Improving Collocated Robot Teleoperation with Augmented Reality," in *Human-Robot Interaction*. ACM/IEEE, 78–86. doi:10.1145/3171221.3171251

Herlant, L. V., Holladay, R. M., and Srinivasa, S. S. (2016). "Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7-10 March 2016, 35–42.

Huang, J., and Cakmak, M. (2017). "Code3: A System for End-To-End Programming of Mobile Manipulator Robots for Novices and Experts," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6-9 March 2017, 453–462.

Jain, S., Farshchiansadegh, A., Broad, A., Abdollahi, F., Mussa-Ivaldi, F., and Argall, B. (2015). Assistive Robotic Manipulation through Shared Autonomy and a Body-Machine Interface. *IEEE Int. Conf. Rehabil. Robot* 2015, 526–531. doi:10.1109/ICORR.2015.7281253

Jain, S., and Argall, B. (2020). Probabilistic Human Intent Recognition for Shared Autonomy in Assistive Robotics. *J. Hum.-Robot Interact.* 9 (1), 1–23. doi:10.1145/3359614

Javdani, S., Admoni, H., Pellegrinelli, S., Srinivasa, S. S., and Bagnell, J. A. (2018). Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming. *Int. J. Robotics Res.* 37 (7), 717–742. doi:10.1177/0278364918776060

Jeon, H. J., Losey, D. P., and Sadigh, D. (2020). "Shared Autonomy with Learned Latent Actions," in *Robotics: Science and Systems*. doi:10.15607/rss.2020.xvi.011

Jia, Y., Xi, N., Liu, S., Wang, Y., Li, X., and Bi, S. (2014). Quality of Teleoperator Adaptive Control for Telerobotic Operations. *Int. J. Robotics Res.* 33 (14), 1765–1781. doi:10.1177/0278364914556124

Kent, D., Saldanha, C., and Chernova, S. (2017). "A Comparison of Remote Robot Teleoperation Interfaces for General Object Manipulation," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6-9 March 2017, 371–379. doi:10.1145/2909824.3020249

Klemmer, S. R., Hartmann, B., and Takayama, L. (2006). "How Bodies Matter," in Proceedings of the 6th ACM conference on Designing Interactive systems - DIS '06, University Park PA USA, June 26 - 28, 2006, 140. doi:10.1145/1142405.1142429

Koh, K. H., Farhan, M., Ching Yeung, K. P., Ho Tang, D. C., Yee Lau, M. P., Cheung, P. K., et al. (2021). Teleoperated Service Robotic System for On-Site

Surface Rust Removal and protection of High-Rise Exterior Gas Pipes. *Automation in Construction* 125, 103609. doi:10.1016/j.autcon.2021.103609

Kolling, A., Nunnally, S., and Lewis, M. (2012). "Towards Human Control of Robot Swarms," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 89–96.

Kortenkamp, D., Bonasso, R. P., Ryan, D., and Schreckenghost, D. (1997). Traded Control with Autonomous Robots as Mixed Initiative Interaction. *AAAI Tech. Rep.* 04, 89–94.

Kristoffersson, A., Coradeschi, S., and Loutfi, A. (2013). A Review of mobile Robotic Telepresence. *Adv. Human-Computer Interaction* 2013, 1–17. doi:10.1155/2013/902316

Labonte, D., Boissy, P., and Michaud, F. (2010). Comparative Analysis of 3-D Robot Teleoperation Interfaces with Novice Users. *IEEE Trans. Syst. Man. Cybern. B* 40 (5), 1331–1342. doi:10.1109/tsmcb.2009.2038357

Lee, M. K., and Takayama, L. (2011). "Now, I Have a Body": Uses and Social Norms for mobile Remote Presence in the Workplace," in *Proceedings of Human Factors in Computing Systems: CHI 2011*. Vancouver, CA, 33–42.

Leeper, A., Hsiao, K., Ciocarlie, M., Takayama, L., and Gossow, D. (2012). "Strategies for Human-In-The-Loop Robotic Grasping," in 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Boston, MA, USA, 5-8 March 2012, 1–8.

Li, Zhi., Moran, P., Dong, Q., Shaw, R. J., and Hauser, K. (2017). "Development of a Tele-Nursing mobile Manipulator for Remote Care-Giving in Quarantine Areas," in 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May-3 June 2017, 3581–3586.

Lin, C. T., Chuang, S. W., Chen, Y. C., Ko, L. W., Liang, S. F., and Jung, T. P. (2007). EEG Effects of Motion Sickness Induced in a Dynamic Virtual Reality Environment. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2007, 3872–3875. doi:10.1109/IEMBS.2007.4353178

Liu, K., Sakamoto, D., Inami, M., Igarashi, T., and Rouge, B. (2011). "Roboshop : Multi-Layered Sketching Interface for Robot Housework Assignment and Management," in *Human Factors in Computing (CHI)*. ACM.

Liu, Y., and Nejat, G. (2013). Robotic Urban Search and rescue: A Survey from the Control Perspective. *J. Intell. Robot Syst.* 72 (2), 147–165. doi:10.1007/s10846-013-9822-x

Mast, M., Materna, Z., Španěl, M., Weisshardt, F., Arbeiter, G., Burmester, M., et al. (2015). Semi-Autonomous Domestic Service Robots: Evaluation of a User Interface for Remote Manipulation and Navigation with Focus on Effects of Stereoscopic Display. *Int. J. Soc. Robotics* 7 (2), 183–202. doi:10.1007/s12369-014-0266-7

Mehr, N., Horowitz, R., and Dragan, A. D. (2016). "Inferring and Assisting with Constraints in Shared Autonomy," in *Conference on Decision and Control, CDC*. IEEE, 6689–6696.

Mingyue Ma, L., Fong, T., Micire, M. J., Kim, Y. K., and Feigh, K. (2018). *Human-Robot Teaming: Concepts and Components for Design*, 649–663. doi:10.1007/978-3-319-67361-5_42

Murphy, R. R., and Tadokoro, S. (2019). *User Interfaces for Human-Robot Interaction in Field Robotics*. Springer International Publishing.

Neustaedter, C., Gina, V., Procyk, J., and Hawkins, D. (2016). To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. *Proc. ACM Conf. Comput. Supported Coop. Work, CSCW* 27, 418–431. doi:10.1145/2818048.2819922

Nicolis, D. (2020). *A General Framework for Shared Control in Robot Teleoperation with Force and Visual Feedback*. Springer Briefs in Applied Sciences and Technology, 119–131. Special Topics in Information Technology. doi:10.1007/978-3-030-32094-2_9

Nielsen, C. W., Goodrich, M. A., and Ricks, R. W. (2007). Ecological Interfaces for Improving mobile Robot Teleoperation. *IEEE Trans. Robot.* 23 (5), 927–941. doi:10.1109/tro.2007.907479

Niemeyer, G., Preusche, C., and Hirzinger, G. (2016). "Telerobotics," in *Springer Handbook of Robotics* (Berlin, Heidelberg: Springer), 741–757. doi:10.1007/978-3-319-32552-1_43

Norman, D. A. (1986). in *Cognitive Engineering*. Editors D. A Norman and S Draper (Erlbaum Associates), 31–62. doi:10.1201/b15703-3

Norton, A., Ober, W., Baraniecki, L., McCann, E., Scholtz, J., Shane, D., et al. (2017). Analysis of Human–Robot Interaction at the DARPA Robotics Challenge Finals. *Int. J. Robot. Res.* 36 (5-7), 483–513.

Norton, A., Ober, W., Baraniecki, L., Shane, D., Skinner, A., and Yanco, H. (2018). "The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue," in *Springer Tracts in Advanced Robotics* Editors M. Spenko, S. Buerger, K. Iagnemma, and D. Shane. (Cham: Springer), 121.

Ochiai, Y., Takemura, K., Ikeda, A., Takamatsu, J., and Ogasawara, T. (2014). "Remote Control System for Multiple mobile Robots Using Touch Panel Interface and Autonomous Mobility," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3272–3277. doi:10.1109/IROS.2014.6943017

Okamura, E., and Tanaka, F. (2016). "A Pilot Study about Remote Teaching by Elderly People to Children over a Two-Way Telepresence Robot System," in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7-10 March 2016, 489–490. doi:10.1109/hri.2016.7451820

Olsen, D. R., and Wood, S. B. (2004). "Fan-out: Measuring Human Control of Multiple Robots," in *Human Factors in Computing Systems (CHI)*. ACM, 231–238.

Peskoe-Yang, L. (2019). *Paris Firefighters Used This Remote-Controlled Robot to Extinguish the Notre Dame Blaze*. IEEE Spectrum. April 22.

Pfeiffer, C., and Scaramuzza, D. (2021). Human-piloted Drone Racing: Visual Processing and Control. *IEEE Robot. Autom. Lett.* 6 (2), 3467–3474. doi:10.1109/lra.2021.3064282

Quigley, M., Goodrich, M. A., and Beard, R. W. (2004). "Semi-autonomous Human-UAV Interfaces for Fixed-wing Mini-UAVs," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) IEEE*, 3, 2457–2462.

Rae, I., and Neustaedter, C. (2017). "Robotic Telepresence at Scale," in *Conference on Human Factors in Computing Systems*. ACM, 313–324. doi:10.1145/3025453.3025855

Rakita, D., Mutlu, B., and Gleicher, M. (2017). "A Motion Retargeting Method for Effective Mimicry-Based Teleoperation of Robot Arms," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6-9 March 2017, 361–370. doi:10.1145/2909824.3020254

Rakita, D., Mutlu, B., and Gleicher, M. (2018). "An Autonomous Dynamic Camera Method for Effective Remote Teleoperation. ACM/IEEE," in *Human-Robot Interaction*, 325–333. doi:10.1145/3171221.3171279

Rakita, D., Mutlu, B., Gleicher, M., and Hiatt, L. M. (2019b). Shared Control-Based Bimanual Robot Manipulation. *Sci. Robot.* 4, 30. doi:10.1126/scirobotics.aaw0955

Rakita, D., Mutlu, B., and Gleicher, M. (2019a). "Remote Telemanipulation with Adapting Viewpoints in Visually Complex Environments," in *Robotics: Science and Systems XV*. Belfast, Ireland. doi:10.15607/rss.2019.xv.068

Rea, D. J., Hanzaki, M. R., Bruce, N., and Young, J. E. (2017a). "Tortoise and the Hare Robot Slow and Steady Almost Wins the Race, but Finishes More Safely," in *Robot and Human Interactive Communication RO-MAN* (IEEE), 1–6. doi:10.1109/roman.2017.8172418

Rea, D. J. (2020). *Now You're Teleoperating with Power: Learning from Video Games to Improve Teleoperation Interfaces*. Winnipeg, Canada: University of Manitoba.

Rea, D. J., Seo, S. H., Bruce, N., and Young, J. E. (2017b). "Movers, Shakers, and Those Who Stand Still: Visual Attention-Grabbing Techniques in Robot Teleoperation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM/IEEE, 398–407.

Rea, D. J., and Young, J. E. (2019a). "Backseat Teleoperator : Affective Feedback with On-Screen Agents to Influence Teleoperation," in Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction, Daegu, Korea (South), 11-14 March 2019, 19–28. doi:10.1109/hri.2019.8673014

Rea, D. J., and Young, J. E. (2018). "It's All in Your Head: Using Priming to Shape an Operator's Perceptions and Behavior during Teleoperation," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5-8 March 2018, 32–40.

Rea, D. J., and Young, J. E. (2019b). "Methods and Effects of Priming a Teloperator's Perception of Robot Capabilities," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 11-14 March 2019, 739–741.

Rea, D. J., Seo, S. H., and Young, J. E. (2020). Social Robotics for Nonsocial Teleoperation: Leveraging Social Techniques to Impact Teleoperator

Performance and Experience. *Curr. Robot Rep.* 1 (4), 287–295. doi:10.1007/s43154-020-00020-7

Reddy, S., Dragan, A. D., and Levine, S. (2018). "Shared Autonomy via Deep Reinforcement Learning," in *Robotics: Science and Systems (RSS)*. New York, NY, United States. doi:10.15607/rss.2018.xiv.005

Reveleau, A., Ferland, F., Labbé, M., Létourneau, D., and Michaud, F. (2015). Visual Representation of Sound Sources and Interaction Forces in a Teleoperation Interface for a Mobile Robot. *J. Human-Robot Interaction* 4 (2), 1. doi:10.5898/jhri.4.2.reveleau

Richer, J., and Drury, J. L. (2006). "A Video Game-Based Framework for Analyzing Human-Robot Interaction: Characterizing Interface Design in Real-Time Interactive Multimedia Applications," in *Human-Robot Interaction.* ACM, 266–273.

Saakes, D., Choudhary, V., Sakamoto, D., Inami, M., and Igarashi, T. (2013). "A Teleoperating Interface for Ground Vehicles Using Autonomous Flying Cameras," in *International Conference on Artificial Reality and Telexistence (ICAT)* Tokyo, Japan: IEEE, 13–19. doi:10.1109/icat.2013.6728900

Sakamoto, D., Honda, K., Inami, M., and Igarashi, T. (2009). "Sketch and Run: a Stroke-Based Interface for home Robots," in *Conference on Human Factors in Computing Systems.* ACM, 197–200.

Sarter, N. B., and Woods, D. D. (1995). How in the World Did We Ever Get into that Mode? Mode Error and Awareness in Supervisory Control. *Hum. Factors* 37 (1), 5–19. doi:10.1518/001872095779049516

Schilling, K., Roth, H., and Lieb, R. (1997). "Teleoperations of Rovers. From Mars to Education," in *International Symposium on Industrial Electronics*. IEEE, 257–262.

Seo, S. H., Rea, D. J., Wiebe, J., and Young, J. E. (2017a). "Monocle: Interactive Detail-In-Context Using Two Pan-And-Tilt Cameras to Improve Teleoperation Effectiveness," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN), 962–967. doi:10.1109/roman.2017.8172419

Seo, S. H., Young, J. E., and Irani, P. (2020). "How Are Your Robot Friends Doing? A Design Exploration of Graphical Techniques Supporting Awareness of Robot Team Members in Teleoperation," in *International Journal of Social Robotics.* Springer.

Seo, S. H., Young, J. E., and Irani, P. (2017b). "Where Are the Robots? In-Feed Embedded Techniques for Visualizing Robot Team Member Locations," in *Robot and Human Interactive Communication*. IEEE, 522–527.

Settimi, A., Pavan, C., Varricchio, V., Ferrati, M., Mingo Hoffman, E., Rocchi, A., et al. (2014). "A Modular Approach for Remote Operation of Humanoid Robots in Search and rescue Scenarios," in *Modelling and Simulation for Autonomous Systems (MESAS)*, 8906. Springer, 200–205. doi:10.1007/978-3-319-13823-7_18

Singh, A., Seo, S. H., Hashish, Y., Nakane, M., Young, J. E., and Bunt, A. (2013). "An Interface for Remote Robotic Manipulator Control that Reduces Task Load and Fatigue," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 738–743. doi:10.1109/roman.2013.6628401

Small, N., Lee, K., and Mann, G. (2018). An Assigned Responsibility System for Robotic Teleoperation Control. *Int. J. Intell. Robot Appl.* 2 (1), 81–97. doi:10.1007/s41315-018-0043-0

Squire, P. N., and Parasuraman, R. (2010). Effects of Automation and Task Load on Task Switching during Human Supervision of Multiple Semi-autonomous Robots in a Dynamic Environment. *Ergonomics* 53 (8), 951–961. doi:10.1080/00140139.2010.489969

S. Tadokoro (Editor) (2019). *Disaster Robotics* (Springer).

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). "Common Metrics for Human-Robot Interaction," in *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*. ACM, 33. doi:10.1145/1121241.1121249

Suarez Fernandez, R. A., Luis Sanchez-Lopez, J., Sampedro, C., Bavle, H., Molina, M., and Campoy, P. (2016). "Natural User Interfaces for Human-Drone Multi-Modal Interaction," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 1013–1022. doi:10.1109/icuas.2016.7502665

Sugiura, Y., Sakamoto, D., Withana, A., Inami, M., and Igarashi, T. (2010). "Cooking with Robots," in Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, Atlanta Georgia USA, April 10 - 15, 2010, 2427. doi:10.1145/1753326.1753693

Szafir, D., and Szafir, D. A. (2021). "Connecting Human-Robot Interaction and Data Visualization," in *ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 281–292. doi:10.1145/3434073.3444683

Teng, W. C., Kuo, Y. C., and Yanu Tara, R. (2013). "A Teleoperation System Utilizing Saliency-Based Visual Attention," in *IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 139–144. doi:10.1109/smc.2013.31

Tsui, K. M., Adam Norton, D. J. B., McCann, E., Medvedev, M. S., and Yanco, H. A. (2013). "Design and Development of Two Generations of Semi-autonomous Social Telepresence Robots," in *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*. IEEE.

Tsui, K. M., Dalphond, J. M., Brooks, D. J., Medvedev, M. S., McCann, E., Allspaw, J., et al. (2015). Accessible Human-Robot Interaction for Telepresence Robots: A Case Study. *Paladyn, J. Behav. Robotics* 6 (1), 1–29. doi:10.1515/pjbr-2015-0001

Turchetti, G., Palla, I., Pierotti, F., and Cuschieri, A. (2012). Economic evaluation of da Vinci-assisted robotic surgery: A systematic review. *Surg. Endosc.* 26 (3), 598–606. doi:10.1007/s00464-011-1936-2

Valiton, A., and Li, Z. (2020). "Perception-Action Coupling in Usage of Telepresence Cameras," in *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, 3846–3852. doi:10.1109/icra40945.2020.9197578

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Matarić, M. J. (2006). "The Role of Physical Embodiment in Human-Robot Interaction," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 117–122. doi:10.1109/roman.2006.314404

Wang, J., and Lewis, M. (2007). "Human Control for Cooperating Robot Teams," in *HRI 2007 - Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*. ACM/IEEE, 9–16. doi:10.1145/1228716.1228719

Wong, C. Y., and Seet, G. (2017). Workload, Awareness and Automation in Multiple-Robot Supervision. *Int. J. Adv. Robotic Syst.* 14 (3), 1–16. doi:10.1177/1729881417710463

Yanco, H. A., Norton, A., Ober, W., Shane, D., Skinner, A., and Vice, J. (2015). Analysis of Human-Robot Interaction at the DARPA Robotics Challenge Trials. *J. Field Robotics* 32 (3), 420–444. doi:10.1002/rob.21568

Yanco, H. A., and Drury, J. (2004b). Classifying Human-Robot Interaction: An Updated Taxonomy. *IEEE Int. Conf. Syst. Man Cybernetics* 3, 2841–2846. doi:10.1109/ICSMC.2004.1400763

Yanco, H. A., Drury, J. L., and Scholtz, J. (2004). Beyond Usability Evaluation : Analysis of Human-Robot Interaction at a Major Robotics Competition. *Human Comput. Inter. ACM* 19–1, 931414160. doi:10.1207/s15327051hci1901&2_6

Yanco, H. A., and Drury, J. (2004a). ""Where Am I?" Acquiring Situation Awareness Using a Remote Robot Platform," in *IEEE Systems, Man and Cybernetics*. IEEE, 2835–2840.

Yang, J., Kamezaki, M., Sato, R., Iwata, H., and Sugano, S. (2015). "Inducement of Visual Attention Using Augmented Reality for Multi-Display Systems in Advanced Tele-Operation," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 Sept.-2 Oct. 2015, 5364–5369.

Young, J. E., Kamiyama, Y., Reichenbach, J., Igarashi, T., and Sharlin, E. (2011). *How to Walk a Robot: A Dog-Leash Human-Robot Interface RO-MAN*. IEEE, 376–382.

Young, S. N., and Peschel, J. M. (2020). Review of Human-Machine Interfaces for Small Unmanned Systems with Robotic Manipulators. *IEEE Trans. Human-mach. Syst.* 50 (2), 131–143. doi:10.1109/thms.2020.2969380

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Robotics and AI

REVIEW
published: 11 April 2022
doi: 10.3389/frobt.2022.719639

Designing Expandable-Structure Robots for Human-Robot Interaction

*Hooman Hedayati[1], Ryo Suzuki[2], Wyatt Rees[1], Daniel Leithinger[1,3] and Daniel Szafir[1,3,4]**

[1]Department of Computer Science, University of Colorado, Boulder, CO, United States, [2]Department of Computer Science, University of Calgary, Calgary, AB, Canada, [3]ATLAS Institute, University of Colorado, Boulder, CO, United States, [4]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

In this paper, we survey the emerging design space of expandable structures in robotics, with a focus on how such structures may improve human-robot interactions. We detail various implementation considerations for researchers seeking to integrate such structures in their own work and describe how expandable structures may lead to novel forms of interaction for a variety of different robots and applications, including structures that enable robots to alter their form to augment or gain entirely new capabilities, such as enhancing manipulation or navigation, structures that improve robot safety, structures that enable new forms of communication, and structures for robot swarms that enable the swarm to change shape both individually and collectively. To illustrate how these considerations may be operationalized, we also present three case studies from our own research in expandable structure robots, sharing our design process and our findings regarding how such structures enable robots to produce novel behaviors that may capture human attention, convey information, mimic emotion, and provide new types of dynamic affordances.

Keywords: deployable robot, human-robot interaction, modular robot, origami robotics, deployable structures, shape-changing robots

OPEN ACCESS

Edited by:
Ginevra Castellano, Uppsala University, Sweden

Reviewed by:
Anaís Garrell, Institute of Robotics and Industrial Computing (CSIC), Spain
Ye Gu, Shenzhen Technology University, China
Amit Kumar Pandey, BeingAI Limited, France

***Correspondence:**
Daniel Szafir
daniel.szafir@cs.unc.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

***Received:** 02 June 2021*
***Accepted:** 15 March 2022*
***Published:** 11 April 2022*

Citation:
Hedayati H, Suzuki R, Rees W, Leithinger D and Szafir D (2022) Designing Expandable-Structure Robots for Human-Robot Interaction. Front. Robot. AI 9:719639. doi: 10.3389/frobt.2022.719639

1 INTRODUCTION

The ability to dynamically change shape and size is a key evolutionary advantage for many biological organisms. For example, pufferfish (Tetraodontidae) and the frilled lizard (*Chlamydosaurus kingii*) can change size as a self-defense mechanism, with the pufferfish able to expand up to three times their original size to warn predators and the frilled lizard able to expand a large frill around its neck, which is folded most of time, when threatened. Other organisms use size and/or shape changes for different purposes. For instance, male magnificent frigatebirds (*Fregata magnificens*) inflate their red throats to attract females, while octopuses change their structures to adapt to dynamic changes in their environment or interact with particular objects. Several fields have adapted this idea and developed shape-changing structures as solutions to various engineering challenges, leading to innovations in the automobile industry (e.g., roof structures in convertible cars), architecture [e.g., temporary exhibition rooms (Escrig and Valcarcel, 1993)], and design (e.g., self-inflating life vests). In addition, human-computer interaction (HCI) researchers have investigated shape-changing properties for developing new types of physical user interfaces (Rasmussen et al., 2012).

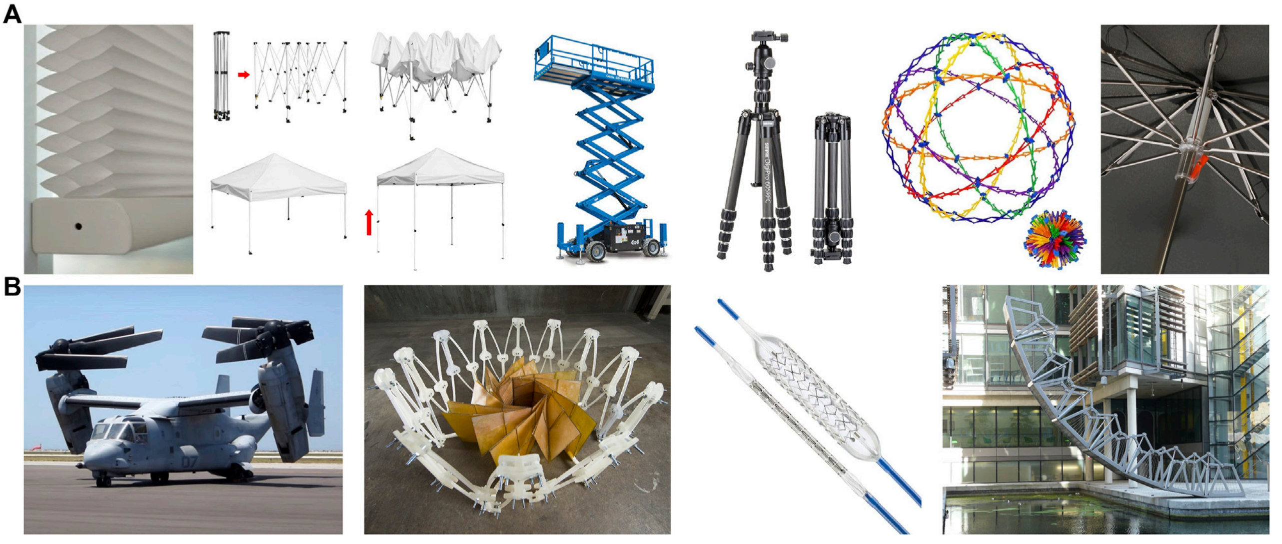


FIGURE 1 | Expandable structures are found in a variety of everyday items, such as window shades, canopies, construction equipment, tripods and stands, toys, and umbrellas **(A)**. Expandable structures are also used in a variety of industrial and scientific purposes, including foldable aircraft, satellite design, medical devices, and architecture **(B)**.

Many of these systems can be described as *expandable structures*: [1] constructions that can change shape, and size using various linkages and joints (Pellegrino, 2002). Many expandable structures can be found in nature, such as in the leaves of hornbeams, flower petals, and the hind wings of beetles (Vincent, 2000; Wang et al., 2017). Engineered expandable structures, a subclass of more general shape-changing technologies, are present in a variety of common consumer products, such as umbrellas, Hoberman spheres, and Origami. Such structures are also used in a diverse set of industrial and scientific equipment, including various forms of construction cranes, stents and other medical devices, foldable satellites, and certain architectural designs, as in adaptive and morphing building structures (Del Grosso and Basso, 2010). **Figure 1** illustrates the diversity of expandable structure, showcasing their applications across everyday and specialized items. There are several methods for changing the size and shape of expandable structures, including mechanical mechanisms (e.g., scissor assemblies, bistable structures, isokinetic/Hoberman mechanisms, etc.), pneumatic or hydraulic mechanisms (e.g., inflatable structures), or through thermal or electrical stimulation of certain materials [e.g., shape memory polymers (Liu et al., 2014)].

In this work, we are primarily interested in expandable structures and related shape-changing technologies in the context of human-robot interaction (HRI) research and applications, including interface technologies, haptics, visualization, and robotics. For instance, one of the primary uses of expandable and shape-changing structures from user interface research has been the development of novel technologies that provide users with physically dynamic interfaces [**Figure 2**, see (Alexander et al., 2018) and (Rasmussen et al., 2012) for full survey of this space]. The goals of such research have strong alignment with many traditional goals of HRI, where shape-changing technologies have been applied to develop devices that can adapt to users and the environment in new ways, communicate information to users, and/or provide novel, adaptive affordances. As an example, researchers have designed multi-touch display surfaces, where each touch point can be deformed to be convex, flat, or concave (Stevenson et al., 2010). This expandable surface matches the physical shape of the display to its visual counterpart, enabling more intuitive interactions, and we can envision HRI researchers applying similar methods to developing novel robot interfaces for teleoperation or supervision. Beyond such physical interfaces, expandable structures have also been used to create brain-computer interfaces (BCIs), which are also being explored for robotics. In (Jiang et al., 2020), expandable fiber probes adapt to contact various parts of nearby brain tissue, enabling scanning of a greater area of brain tissue with fewer surgical insertions and reducing patient risks in such procedures.

Another major focus of shape-changing structure research relevant to human interaction has been developing new forms of haptic feedback interfaces, particularly for use with Virtual Reality (VR). This research generally leverages expandable structures to provide *encountered-type haptics*, in which certain aspects of the surrounding real-world environment shift dynamically to provide physically resistive forces when users make contact with virtual objects. For example, FEELEX (Iwata et al., 2001) and shapeShift (Abtahi and Follmer, 2018; Siu

[1]Such structures are also commonly described as "deployable." As we are primarily focused on the use of such structures with robotics to improve HRI, in this paper we use the terminology of "expandable" to avoid potential confusion with the notion of "deploying" robots for particular applications.

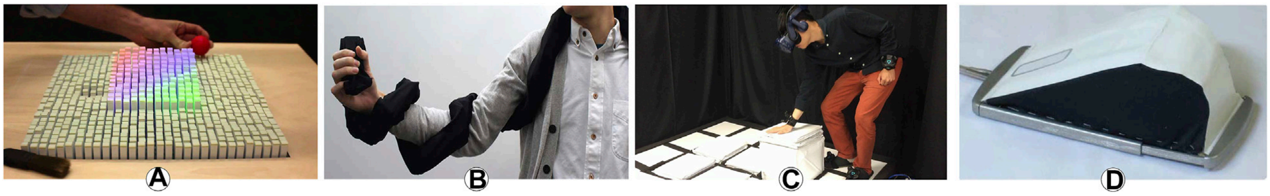


FIGURE 2 | Examples of various ways shape-changing interfaces researchers and designers have proposed to augment human-computer interaction: **(A)** dynamically actuated shape displays such as Materiable (Nakagaki et al., 2016b), **(B)** deformable, actuated linkages such as LineFORM (Nakagaki et al., 2015), **(C)** inflatable structures such as TilePoP (Teng et al., 2019), and **(D)** Inflatable Mouse: a volume-adjustable mouse with air-pressure-sensitive input (Kim et al., 2008).

et al., 2018) implemented dynamically actuated shape displays using an array of actuators in combination with a flexible screen and electro-active polymers, respectively. These devices provide the capability to simulate varying surfaces and shapes in VR. At a larger scale, TilePoP (Teng et al., 2019) and LiftTiles (Suzuki et al., 2020b) investigated the use of inflatable actuators to dynamically change the user's surrounding physical environment to provide haptic proxies. Each of these haptic displays utilize expandable structures to create a VR experience that users perceive as more realistic, as the visual cues associated with virtual object interaction can be accompanied by real forces.

Recent work has also shown the potential of expandable structures for several other interaction-focused applications, such as visualization, design, and education. For example, the HATs (Mi and Sugimoto, 2011) and G-Raff (Kim and Nam, 2015) systems used height-changing structures to synchronize the height of objects with digital content on a tabletop surface, enabling intuitive interaction with 3D spacial data. LineFORM (Nakagaki et al., 2015) demonstrated how a physical line made from actuated linkages can transform into a wristwatch, a phone, and several other objects. This type of dynamic physical display allows for richer interactions with a wide array of objects and data. Additionally, it can provide new constraints on user interactions in order to provide guidance, presenting opportunities for the device to scaffold learning. Moreover, highly extendable linear actuators can achieve both shape- and size-changing transformations (Takei et al., 2012; Hammond et al., 2017; Hawkes et al., 2017). These devices present opportunities in several domains, ranging from enabling dynamic and self-erecting architecture to providing increased mobility in search and rescue operations by supporting adaptation to irregular terrain. The Topobo (Raffle et al., 2004) and ShapeClip (Hardy et al., 2015) structures allow a designer to construct different geometries of shape-changing interfaces and have shown potential for enhancing early education by helping children learn about relationships between physical formations and physical properties, such as balance and leverage.

In this paper, we focus on the integration of expandable structures and robotics as a promising avenue for improving human-robot interaction (HRI). In recent years, researchers and engineers have leveraged expandable structures for several robotic applications (Felton et al., 2014; Kornatowski et al., 2017; Perez-Guagnelli et al., 2018). For example, expandable structures have helped aerial robots navigate through narrow spaces (Falanga et al., 2018) and robot arms reach confined areas (Shikari and Asada, 2018). However, such prior deployments of expandable structures for robotics have primarily focused on specific aspects of robot task and/or control (e.g., manipulation, locomotion, etc.). In this paper, we instead categorize a broad range of HRI-relevant factors and implementation considerations while synthesizing several interaction-based use-cases for expandable structures and shape-changing robots. We use these categorizations as part of detailing an incipient design space in how such technologies may improve robot interactions with collocated humans.

As examples of this design space, we also highlight three specific implementations of expandable structures for HRI from our own research, including *RoomShift*, a ground robot that uses expandable structures to move furniture in a room in order to provide haptics for a human working in virtual reality (Suzuki et al., 2020a), *PufferBot*, an expandable structure for aerial robots that can improve safety while also introducing a new signaling mechanism to communicate with nearby humans (Hedayati et al., 2020), and *ShapeBot*, a miniature tabletop robot that can change shape individually and also as part of a larger ShapeBot swarm to convey various information to users (Suzuki et al., 2019b).

2 EXPANDABLE STRUCTURES FOR ROBOTICS

To date, there has been very little work exploring robots with expandable structures from a human-robot interaction perspective. Instead, most prior work has focused on the mechanical aspects of building expandable structure robots, which fall within a broader category of shape-changing robots. While precisely classifying the full space of shape-changing robots is challenging, as some robots might cross categorical boundaries, systems developed in prior research generally fall into one of the following major groups: **modular** self-reconfigurable robots, **origami**-like robots, **tensegrity** robots, **soft** robots, or deployable/**expandable** robots.

Modular robots are robots made of identical or similar elements that can be attached in different ways to form different group structures (Støy, 2015; Shang et al., 2018). The Reconfigurable Modular Manipulator System (RMMS) (Kelmar and Khosla, 1990) was one of the earliest modular robots and consisted of a set of modular links and joints of various sizes that could be reconfigured according to specific tasks. This system introduced dynamism to industrial robotics, enabling faster productivity and reduced costs. In a similar vein, a modular isomorphic master-slave robotic system was developed (Zheng et al., 2013) to enable master robots to be highly adaptable to varying structures and degrees of freedom in slave robots, further increasing productivity and reducing costs in a wider range of domains. Chain-type robots, such as PolyPod (Yim, 1994), CONRO (Castano et al., 2000), and PolyBot (Duff et al., 2001), are robots constructed as a connected series of modular parts, simplifying and expanding the level of dynamic abilities that robots can achieve. Lattice-type robots, such as Molecule (Kotay et al., 1998) and 3-D-Unit (Murata et al., 1998), are constructed as a grid-like network structure of modular pieces. These robots provide similar benefits as chain-type robots, while also expanding their reconfigurability into another dimension in space. Hybrid modular robots, such as M-TRAN III (Kurokawa et al., 2008) and SuperBot (Salemi et al., 2006), use a combination of chain- and lattice-type structures. Changibles (Roudaut et al., 2014) and Cubimorph (Roudaut et al., 2016) are shape-changing robots that leverage a modular and reconfigurable design to achieve different geometries, allowing for richer and more intuitive interactions with dynamic shape displays. ChainFORM (Nakagaki et al., 2016a) integrates modular sensing, display, and actuation to further enhance interactions.

Another category of robots that exhibit shape- and/or size-changing properties are Origami-like robots. Origami has been used in many engineering areas (Okuzaki et al., 2008; Ma and You, 2013) and is increasingly feasible for robotics due to improvements in fabrication and actuator technologies. Examples of origami-like robots include robotic sheets that can be folded into different morphologies (Hawkes et al., 2010) and a set of programmable triangles which can create different patterns (Belke and Paik, 2017). Origami robots offer several advantages, including the elimination of redundant materials used in separate tasks, reducing the amount of materials needed overall, and their foldable designs may often serve dual purposes, such as providing a robot chassis with built-in protection [e.g., as in origami-inspired mechanisms for aerial robots (Kornatowski et al., 2017; Sareh et al., 2018; Shu and Chirarattananon, 2019)]. To date, most research on Origami robots has focused on physical design and actuation (Lee et al., 2013; Onal et al., 2014; Vander Hoff et al., 2014; Miyashita et al., 2015) or on using smart materials to create self-folding robots (Paik et al., 2010; Paik and Wood, 2012; Tolley et al., 2014; Firouzeh and Paik, 2015). Recently, researchers have also explored Kirigami structures, an extension of Orgami that supports cutting in addition to folding, for deployable robot design (Sedal et al., 2020).

Tensegrity robots and soft robotics take a different approach towards developing actuated systems. Tensegrity robots focus on designing systems made of tensegrity structures (Snelson, 1965), which is an abbreviation of *tensile integrity*. Tensegrity robots are typically formed from constructions of ropes, tube, springs, and joints that provide strength and compliance while being lightweight. As a result, tensegrity robots have particular relevance to space robotics (SunSpiral et al., 2013; Sabelhaus et al., 2015). Currently, most research in tensegrity robotics is focused on design, locomotion, and control (Caluwaerts et al., 2014; Sabelhaus et al., 2015; Zhang et al., 2017; Vespignani et al., 2018; Wang et al., 2019). To the best of our knowledge, such structures have yet to be explored from a human-robot interaction perspective.

In contrast to rigid systems, soft robots actuate elastic and compliant materials, such as rubbers, hydrogels, and silicone elastomers (Coyle et al., 2018). There are a variety of actuation methods for soft robots, including pneumatic, electroactive polymer, tendon driven, shape memory alloy, and electro- and magneto-rheological materials (Das and Nabi, 2019). From a HRI perspective, soft robots may improve safety in collocated use cases due to their complaint nature and have been explored for several applications, including wearable robots that provide human movement assistance (Maeder-York et al., 2014; Park et al., 2014) or convey emotions (Hu and Hoffman, 2019).

In this paper, we are particularly focused on a subclass of shape-changing robots: expandable (i.e., deployable) structure robots that use rigid mechanisms to change their size and shape to improve mobility or gain new interactive capabilities. In terms of mobility, various "reconfigurable" or "hybrid" ground-mobile robots have been developed that may change form to use either wheel or leg locomotion to adjust to changes in terrain [e.g., (Ding and Xu, 2009; Chen et al., 2013; Reid et al., 2020); for a survey, see (Russo and Ceccarelli, 2020)]. Alternatively, the AmphiHex-I presents a design for an amphibious robot with leg-flipper composite propulsion, enabling the robot to walk and move under water (Liang et al., 2012). Such concepts have also been explored in aerial systems, where researchers have created foldable drone frames to enable navigation through confined spaces (Falanga et al., 2018) and hybrid systems, such as HeritageBot, capable of walking and flying (Ceccarelli et al., 2018). Beyond mobility, researchers have used expandable structures for robots in various ways to enable dynamic robot re-sizing. For example, expandable structures have led to deformable wheels for robots (Lee et al., 2013), robots capable of self-folding from a sheet to a 3D structure (Miyashita et al., 2015), and robots arms able to extend to gain additional manipulation reach (Shikari and Asada, 2018). While promising, such research typically details the design of one particular expandable structure robot or application. To help researchers seeking to explore expandable structures for HRI, below we synthesize several implementation considerations for developing expandable structures specifically within the context of robotics and describe a broader design space regarding how expandable structures may afford new methods of interaction between humans and robots.

3 IMPLEMENTATION CONSIDERATIONS

In this section, we present an overview of various implementation details necessary for developing expandable structures for HRI research. To help future researchers and designers better reason

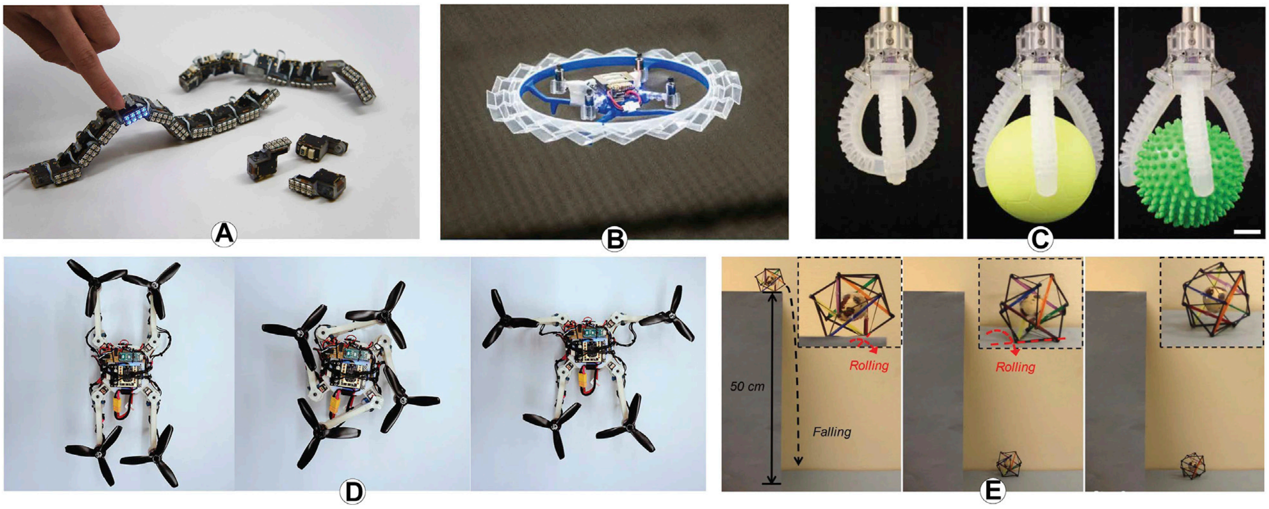


FIGURE 3 | Examples of robots with shape-changing technologies, including **(A)** modular robots in ChainFORM (Nakagaki et al., 2016a), **(B)** origami structures in Rotorigami (Sareh et al., 2018), **(C)** soft materials (Truby et al., 2018), **(D)** deployable structures that enable folding, expansion, and contraction (Falanga et al., 2018), and **(E)** tensegrity systems (Wang et al., 2019).

through the various alternatives and opportunities available for developing expandable structure robotics, we describe considerations involving expandable structure type, actuation method, and integration with robot platforms. **Figure 16** provides a visual reference for these implementation considerations and potential interaction goals, which are detailed in **Section 4**, and shows the decisions we made in each of our case studies described in **Section 5**.

3.1 Expandable Structure Types

Depending on the purpose of the device, expandable structures may use various methods to change their shape and size. While several methods for classifying expandable structures have been proposed [see (Fenci and Currie, 2017) for a survey], we are primarily interested in two main categories: those that utilize mechanical joints and those that utilize the physical properties of continuous materials. One of the most widely-used mechanical methods of expansion is a scissor-like structure [**Figure 4A**, see (Zhao et al., 2009) for a review of the mechanics underlying scissor structures]. Most commonly, these structures allow for linear expansion and retraction, an example of which is the electric scissor lift. However, scissor-like structures may also be used to expand in a radial fashion. Another common structure used for expansion is the Hoberman linkage (**Figure 4B**). This structure is comprised of a similar series of parts as the scissor-like structure, but instead allows for radial expansion. Six Hoberman linkages may be aligned according to the edges of an icosidodecahedron and actuated simultaneously in order to create a Hoberman sphere. Another type of expandable structure that is common among consumer products are those that use retractable plates, such as a camera shutter or movable form of wheelchair ramp used on buses to provide for wheelchair access. A similar concept is found in telescopic structures, which use concentric tubular sections that slide into one another (See **Figure 4D**). Another class of mechanical expandable structures are reel-based structures (Hammond et al., 2017; Suzuki et al., 2019a; Suzuki et al., 2021) or those that use revolute joints to unravel chain- or lattice-type structures (**Figure 4C**). These structures are unique in that they can allow for expansion in all three dimensions of space.

Expandable structures that utilize the physical properties of continuous materials do so with soft, flexible materials, such as silicone (**Figure 4D**). The benefit of these structures typically resides in their ability to take on many different types of shapes and curvatures. Typically, these structures form a 3D surface, which may be expanded and/or morphed into different shapes. For example, one study used a self-expanding silicone stent to help patients with esophageal cancer swallow food (Siddiqui et al., 2007). An example of shape-changing soft expandable structures from user interaction research is PneUI (Yao et al., 2013), which uses soft composite materials to create a shape-changing interface.

When designing an expandable structure, one must carefully analyze the physical domain in which the structure serves a purpose: How many dimensions does the structure need to expand into? How large must the structure be? How strong or rigid does the structure need to be? The answer to these questions will be a primary determining factor in deciding the type of structure that is best suited for the problem. For example, if a structure only needs to expand in one direction and must interact with heavy objects, a rigid, scissor-like structure is a natural choice. On the other hand, if the structure is intended to represent data in various forms or is intended to be touched and deformed by a human, a soft, shape-changing structure may be better suited.

3.2 Actuation Methods

There are several methods researchers may choose to actuate expandable structures, including hydraulic, pneumatic, electric,

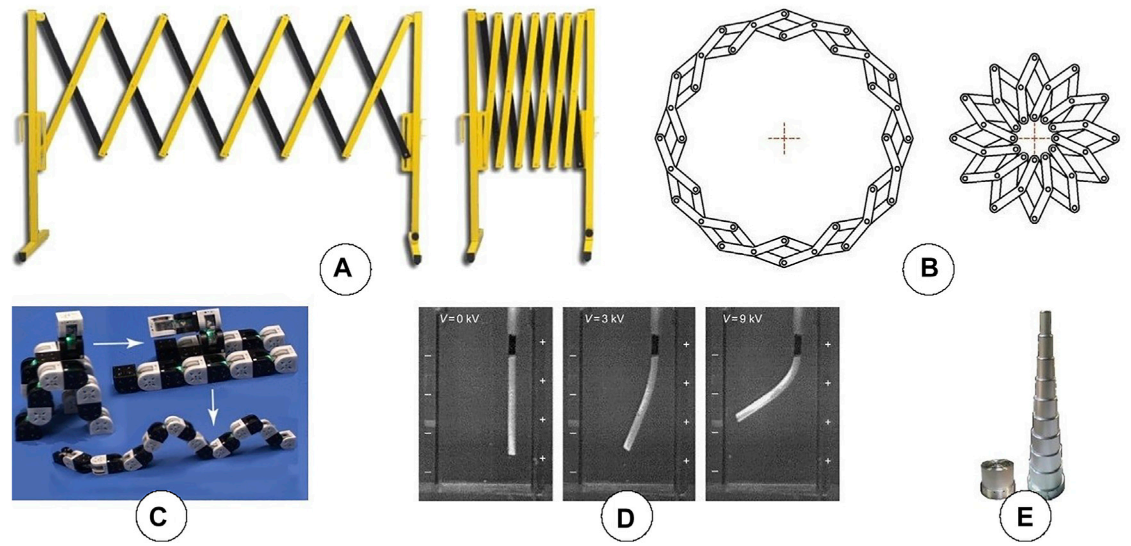


FIGURE 4 | (A) A scissor-like expandable structure, capable of expanding linearly. **(B)** A Hoberman linkage, capable of expanding radially. **(C)** A chain-type expandable structure, capable of expanding various components in all directions in space (Kurokawa et al., 2008). **(D)** An electro-active polymer, which may be capable of forming different shapes. **(E)** A telescopic structure.

and mechanical. Hydraulic actuators consist of a hollow cylindrical tube along which a piston can slide. A hydraulic pump delivers a regulated flow of compressed liquid to move the piston. These actuators are capable of exerting forces of relatively high magnitude, but cannot achieve high acceleration compared to other actuators. Pneumatic actuators work in a similar fashion as hydraulic actuators, but instead use air pressure to move the piston. They can also provide forces of high magnitude with relatively small volumes of air, but require complex systems of components (compressors, reservoirs, filters, etc.) that may result in inefficient energy loss. Electric actuators typically convert the rotational force of an electric rotary motor into a linear movement, which can be done with hydraulic or mechanical mechanisms. Another form of electric actuators use a series of oppositely aligned magnets and electric coils driven in opposing phases to generate a linear force without extraneous mechanical or hydraulic components. Another type of electric actuation uses electro-active polymers, which act like artificial muscles. When an electric current is supplied through the polymer, it contracts (Novack et al., 2021). Releasing the current allows for the polymer to expand again. In this case, the actuation method may act as the expandable structure itself. Mechanical actuators convert rotational force into linear force through the use of components such as belts, screws, or gears. What makes the mechanical actuators different from electrical actuators is that, in mechanical actuators, the energy needed for actuation is stored in a non-electric way, such as in springs.

Different actuation methods provide trade-offs for researchers and designers seeking to create expandable structures for human interaction. For example, if the intended interaction may involve direct physical contact with humans, an actuation method that exerts relatively lower magnitudes of force may enhance user safety—in the case of a system malfunction, there is less potential for harm to the user. Conversely, if an expandable structure is used to alter or manipulate objects and/or the environment, as is the case with structures that enable encountered-type haptics, an actuation method capable of exerting higher magnitudes of force may be necessary. If the expandable structure is intended for visualization of various data, actuation precision or speed may be primary considerations.

3.3 Robot Integration

Integrating expandable structures with robots may require specific considerations of robot type, size, and capabilities. There are several different ways of classifying robots, such as considering morphology (e.g., anthropomorphic/human-like, biological/zoomorphic, or functional) (Fong et al., 2003; Li et al., 2010), capability (e.g., fixed-based manipulation, ground-mobile, ground-mobile manipulators, aerial), or degree of autonomy (Szafir et al., 2017). For traditional and pre-existing robot platforms, expandable structures may be added as sub-components [e.g., an expandable structure for compliant robot grippers as in (Kaur and Kim, 2019)] or as entire frames [e.g., a protective frame around a drone (Hedayati et al., 2020)]. Alternatively, new robots may be designed to leverage expandable structures as central components of the robot itself, as in the Triple Scissor Extender Robot Arm (Shikari and Asada, 2018), a new design for an expandable structure robot arm that supports manipulation in cluttered and confined areas. In both contexts, relevant considerations for roboticisits include power, weight, and structure materials. Power for expandable structures may be self-contained or draw on a central robot power supply, while weight and materials may be selected based on platform needs and application goals. For example, a structure made for manipulation or lifting of heavy objects would require a strong, rigid structure, while a structure made to reduce the impact of collisions would require a more compliant material to reduce the impact force. Prior work has explored expandable structures constructed with various materials, including metal (Shikari and Asada, 2018), ionic polymer-metal composite (Niu et al., 2015), soft silicon (Takei et al., 2011a), latex (Stevenson et al., 2010), plastics (Sedal et al., 2020), and other soft materials. Other promising materials that have yet to be extensively explored for expandable structures integrated with robots include wire structures, wood, and linen. Roboticists seeking to integrate expandable structures with existing platforms must additionally consider how to mount expandable structures in a manner that does not impede robot mobility or existing capabilities while ensuring visibility, potentially by leveraging universal mounting systems or developing custom mounting plates, as in (Suzuki et al., 2020a; Hedayati et al., 2020).

Across all types of robots with expandable structures, researchers must also consider size and expandable structure capabilities. The size of the expandable structure will likely increase with the size of the robot. As the size of the structure increases, so will its weight. With these correlations, structure material may be the primary consideration (i.e., robots with limited payload capacity must make use of lightweight materials for their expandable structures). A similar comparison can be made with smaller robots, which may only be able to support smaller payloads due to mounting challenges. Overall, researchers and developers will need to consider the trade-off space between weight and strength (resulting from material choice and structure design) and payload capacity. Additionally, larger robots that utilize expandable structures as a frame may need additional support mechanisms for the structure to prevent it from collapsing.

In addition to material choice, the expandable structure's intended capabilities will have a large influence on the proper actuation choice for the structure. For example, if the purpose of the expandable structure is to enable better manipulation of potentially heavy objects, the actuation will need to be able to output a large force. In this case, a hydraulic or pneumatic actuator will likely be a good choice. In some special cases, an expandable structure that is expected to endure high forces may be able to use less powerful actuators. For example, if the structure is intended to protect the robot from collisions, one could rely on other mechanisms besides the actuators to prevent the structure from collapsing upon collision. One such mechanism could be pieces of the structure that lock in to place upon expansion of the structure, much like locking one's knees when fully straightening one's legs. On the other hand, if the purpose of the structure is to enhance fine manipulation or to visualize data, a more precise actuation method (e.g. electromechanical) may be required.

4 INTERACTION DESIGN SPACE

In this section, we describe the design space regarding how expandable structures may be integrated with robotic systems to improve human-robot interaction (**Figure 5**). We highlight how expandable structures may provide robots with new ways to interact with both their surrounding environment and collocated humans, expand robots abilities to signal and convey information to humans, improve human-robot safety, and affect experiential aspects of interactions, such as altering aesthetics or enhancing enjoyment, curiosity, or playfulness.

4.1 Adaptive Affordances

Providing robots with the ability to adapt their shape and size based on interaction context opens up many new possibilities in how robots may interact with humans, objects, and their environment. Such adaptation may be related to a specific HRI *task*, where, for example, expandable structures may afford a collaborative or teleoperated robot with new capabilities for manipulating or assembling objects (e.g., altering leverage to adjust objects that would otherwise be too unwieldy or expanding to grasp otherwise out-of-reach objects), new ways to navigate through confined environments that would otherwise be infeasible to operate within, or new ways for multiple robots to work together by combining expandable structures in a team fashion, making use of the fact that many types of expandable structures are modular in design. Alternatively, adaptation may be related to the *user*, where, from a human's point of view, expandable structures might change shape and/or size to indicate different possibilities for user interaction (e.g., a robot that detects an internal fault might change shape to enable a technician easier access to internal components that an expandable structure would guard in normal circumstances). Certain applications may involve adaptation to both task and user, as in the design of expandable structure robotic exoskeletons [e.g., (Li et al., 2019)], where expandable structures may provide singularity-free joints for wearable robots that do not compromise human limb function (Castro et al., 2019). Expandable structures also offer new capabilities for user control of robots, particularly for novices, who may lack the situational awareness or experience necessary to accurately control complex systems such as redundant manipulators or aerial robots leading to self-collisions, crashes, and/or damage to surrounding objects or the environment. Expandable structures may offer a new way in which robot operators may physically "probe" the surrounding robot environment in a safe manner by bumping into other objects, walls, ceilings, etc. without damage. Such an interaction may be used in educational or training scenarios, where users gain confidence and abilities controlling new robotic systems or in real systems, potentially combined with haptic feedback controls, to enhance user awareness of the robot environment. In a similar fashion, expandable structures can enable robots to work with users in new environments that were previously too cluttered or confined (Shikari and Asada, 2018; Hedayati et al., 2020).

One particularly promising application of how expandable structures may provide robots with adaptive, task-based capabilities to support human interaction is through using robots as novel haptic interfaces, especially in conjunction with virtual reality. As described in **Section 2**, encountered-type haptics focuses on providing users with physically resistive forces that simulate virtual objects to enhance user presence in VR. With the ability to change size and shape, a single robot might be able to represent several different sizes or types of objects in a virtual environment. For example, a VR user could interact with several virtual balls of different sizes that are all physically represented by a single robot with a Hoberman sphere structure that expands or contracts to match the size of the ball used at any given time. We detail our own work at the

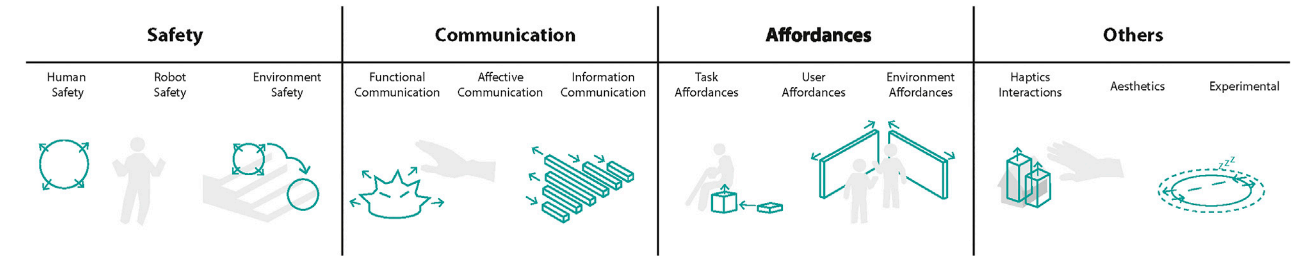


FIGURE 5 | Expandable structures open up new design spaces for human-robot interactions by promoting physical safety, providing novel communicative channels, supporting adaptive affordances, and altering experiential aspects such as aesthetics or enjoyment.

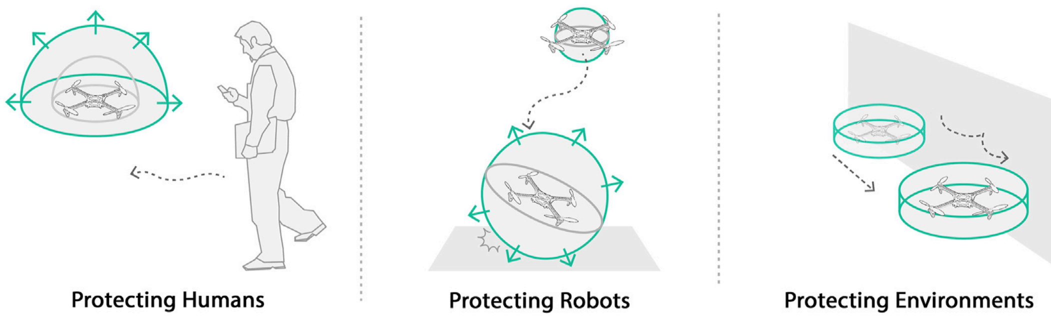


FIGURE 6 | Robots with expandable structures can improve the safety of collocated humans, the robot itself, and the environment.

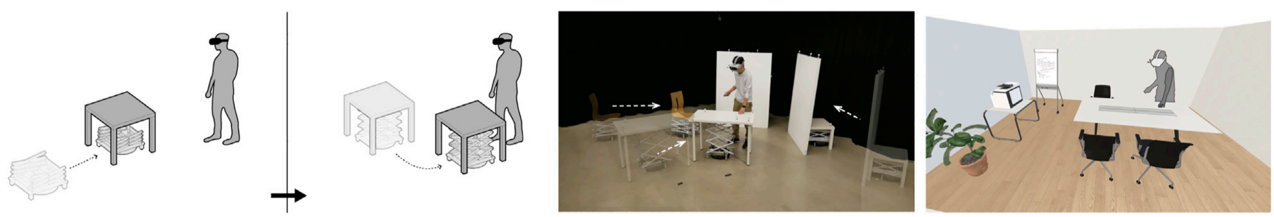


FIGURE 7 | RoomShift is comprised of a swarm of shape-changing robots that provide haptic feedback in VR by manipulating physical furniture and walls. Multiple robots move environmental objects to collectively construct and adapt a physical haptic environment that matches virtual scenes. Above, we show the physical environment with a corresponding virtual scene, with a human silhouette added for a reference.

intersection of expandable structure robots and encountered-type haptics for VR in **Section 5.1**.

4.2 Non-Verbal Communication

One of the ways that we can improve human-robot interaction is by expanding the communication mechanisms available for exchanging information between humans and robots. Prior research has explored a variety of communicative channels, including gaze (Mulu, 2006; Andrist et al., 2012; Andrist et al., 2014; Admoni, 2016; Oliveira et al., 2018), implicit motion (Dragan et al., 2013; Szafir et al., 2014; Sadigh et al., 2016; Zhou et al., 2017; Kwon et al., 2018), gesture (Waldherr et al., 2000), sound (Cha and Matarić, 2016; Cha et al., 2018a), visual displays, lights (Szafir et al., 2015; Baraka et al., 2016; Song and Yamada, 2018), haptics (Guerreiro et al., 2019; Guinness et al., 2019), projection (Pierce et al., 2012; Cauchard et al., 2019), and augmented reality (Hedayati et al., 2018; Walker et al., 2018; Cao et al., 2019; Szafir, 2019; Walker et al., 2019). Expandable structures represent a promising new signalling medium to add to this collection of methods for supporting human-robot information exchange, which may take the form of functional cues regarding the robot and task or affective signals that communicate emotional information.

4.2.1 Functional Communication

We envision that expandable structures may be used to convey a variety of common functional signals to collocated users, including information about internal states (e.g., expansion of the structure might correlate with battery level), higher-level information about processes or tasks (e.g., the percentage of completion of a task), or goals and intent (e.g., expanding a structure the direction the robot intends to move and contracting a structure on the opposite side). Expandable structures might also enhance other methods of signaling in HRI by increasing or decreasing the size of other signaling devices. For example, LED arrays or strips can be flashed in different patterns to signal information to humans. However, the discrenability of such visual signals depends on the distance between the robot and the user and the distance between individual lights. By using an expandable structure, the visual signalling mechanisms could dynamically change their separation, for example contracting to aggregate various disparate visual signals into a cohesive display or separating to make it easier to distinguish individual visual channels from greater distances. To date, we have yet to see any work in HRI examining the use of expandable structures for such functional signalling and we believe it to be a rich, untapped area for future exploration.

4.2.2 Affective Communication

Expandable structure designs are often inspired by plants and animals that increase or decrease their size as a way to escape or frighten predators. Robots equipped with expandable structures might use similar life-like patterns of expansion or contraction to convey affective information. Research has found that affective communication and giving a human-like character can improve human-robot interaction as users perceive the robot to be more intelligent (Admoni, 2016; Cha et al., 2018b). Affective communication conveys the emotions of a social robot (Hu and Hoffman, 2019) or makes information like the robot's intent more understandable to users by exaggerating animations (Szafir et al., 2015). For example, a robot with an expandable structure that could present a danger to collocated humans or is engaged in a critical/uninterruptible task might mimic the example of the pufferfish, which increases in size when feeling threatened, by expanding to warn humans against approaching or coming near the robot. Alternatively, there are some animals that contract as a defense mechanism. For example, the leaves of the shameplant (*Mimosa Pudica*) fold inward and droop when touched or shaken, defending themselves from harm, and re-open a few minutes later. This mechanism offers an alternative inspiration for expandable structure behavior, where a similar contracting mechanism might convey weakness and robot fragility, contrasting the dangerous and intimidating nature of an expansion behavior.

Finally, expandable structure designs can also be inspired by human nature. For example, when people become anxious or afraid, their heart rate increases and they may start breathing faster. These physiological responses are a sign of discomfort and something humans may intuitively understand and feel empathy for. Alternatively, other patterns of behavior (e.g., foot tapping, skipping, etc.) are commonly associated with a variety of other affective states (e.g., irritation, joy, etc.), providing a rich area of inspiration from which HRI researchers may draw. While we believe that expandable structures hold significant promise in conveying affective information in HRI contexts, to date we have yet to see research investigating this space. We discuss our own preliminary investigations in this area in **Section 5.2**.

4.2.3 Data Visualization

Expandable structures may also enable robots to provide new ways of visually communicating data to users. As an example, robots might use linear or radial expansion to physicalize data (e.g., forming physical bar graphs or scatterplots). One advantage that such robots may offer over static data physicalization techniques is the ability to dynamically represent data. In addition, expandable structure robots may also act as a dynamic physical displays, supporting real-time transformations of data into different representations such as bar graphs, line charts, or star graphs. We explore these aspects in **Section 5.3**.

4.3 Safety

Within the broader area of HRI, the sub-field of physical human-robot interaction (pHRI) focuses on concerns related to human safety. Several methods of ensuring physical safety have been identified in the pHRI literature, including safety through control, motion planning, prediction, and consideration of human psychological factors [see (Lasota et al., 2017) for a survey]. Specifically, with regard to physical safety, research has predominantly focused on different methods for managing collisions. Currently, most large robots that are potentially fatal to humans on collision operate only in safety cages. Other, less powerful but still potentially hazardous robots may leverage expandable structures as another form of a physical barrier. Expandable structures provide a simple, yet effective mechanism for creating dynamic boundaries around dangerous and fragile components of a robot. These structures have the potential to provide safety to three different components of any human-robot interaction scenario: the human, the robot itself, and the surrounding environment.

Many robots are comprised of various components that posses large momenta, which can result in a large impact force or pressure upon coming into contact with a human. Expandable structures provide a unique mechanism for physically separating these specific components without imposing large restrictions on robot movement or functionality. In addition to preventing collisions entirely, expandable structures also have some degree of compliance, enabling them to act as an airbag or bumper in order to reduce the impact of any collisions with a human.

Robots may also be dangerous to themselves. Certain components of robots may be fragile, such as drone propellers, or require precise and time-consuming calibration, as is the case in many industrial robots. In such cases, collisions may damage parts or shift components, requiring component replacement or recalibration. For example, the propellers of aerial robots are often extremely fragile. If a propeller comes in to contact with a surrounding object during flight, it is likely to break or deform, resulting in unstable and unpredictable flight patterns. While a static cage (i.e., propeller guard) can provide one way of protecting propellers, it permanently increases the size and shape of the robot, potentially reducing its mobility. In contrast, an expandable structure may expand to protect the propellers when the robot is more likely to collide with surrounding objects (i.e., when flying in constrained areas), and retract when not needed to give the robot more mobility.

In a similar fashion, expandable structures may reduce potential damage to any objects in the surrounding environment. For example, the compliance of expandable structures will mitigate and forces transferred from a robot to any object it hits (walls, instruments, other robots, etc.) in the event of a collision.

4.4 Aesthetic and Experiential Purposes

In addition to the interactive possibilities described above, expandable structures may be integrated with robotics purely for aesthetic purposes, for fun and entertainment, or to simulate users and enhance user enjoyment during the experience of working with robots. For such use, roboticists may take inspiration from aesthetic use of expandable structures in fashion (e.g., smart adaptive garments), art, or architecture. In addition, such structures might be used to give robots additional lifelike traits or quirks, such as enabling robots to mimic expansion and contraction in biological breathing movements in a manner similar to the generation of natural robot motions (Koditschek, 1984) and gaze patterns (Yoshikawa et al., 2006).

5 CASE STUDIES

To advance our vision for how expandable structures may enhance HRI, in this section we detail three of our own research projects integrating expandable structures and robotics within interactive scenarios. We focus on illustrating a broad swath of the design space (e.g., different structures, robots, applications, etc.), showcasing our design and implementation process, and highlighting human responses to such robots. First, we introduce RoomShift (Suzuki et al., 2020a), a large ground robot that uses scissor-like expandable structures to move furniture in a room to enable encountered-type haptics for a human using a virtual reality headset. Next, we describe PufferBot (Hedayati et al., 2020), a medium-size aerial robot with an isokinetic expandable structure that can take several forms and afford three types of expanding behaviors. Finally, we detail ShapeBot (Suzuki et al., 2019b), a miniature tabletop robot that can alter its shape individually and as part of a larger ShapeBot swarm for a variety of purposes, including information visualization and environment manipulation.

5.1 RoomShift

RoomShift (Suzuki et al., 2020a) is a room-scale swarm of off-the-shelf ground robots to which we added large expandable structures to provide the robots with new environment manipulation capabilities. We then leveraged these robots to generate a new haptic feedback mechanism for virtual reality, whereby RoomShift robots reconfigure the physical environment in real time to match various virtual scenes, inspired by shelf-moving robots in robotic warehouses (Guizzo, 2008; Wurman et al., 2008).

5.1.1 Design and Implementation

In their original form, each robot (a Roomba) lacks the capability to manipulate large objects. We added a mechanical lift expandable structure that can extend from 30 to 100 cm, affording the robots the ability to pick up, carry, and place objects such as chairs, tables, and movable walls. When combined with a virtual environment, the RoomShift system enables users to touch, sit, place, and lean against objects in the virtual world. In our current deployment, we have synchronized VR scenes with a 10 m × 10 m physical environment outfitted with an optical motion tracking system to support software that tracks and controls the robots. To do so, we implemented customized software in Unity which gets the user and furniture positions from the motion tracking cameras, creates the VR scene, and compiles the user's commands to control the robots' movement of the furniture. This system continuously maps virtual touchable surfaces in the proximity of users and coordinates the robot swarm to move physical objects to their target locations without colliding with each other or the users. The user and robots do not interact with each other directly. Since the user is fully immersed in the virtual environment, they can only see and interact with the items rendered in the VR scene (e.g., chairs, desks, etc.), which does not include the robots.

In designing RoomShift, we considered and tested several expandable structures and actuation mechanisms, including a pneumatically-actuated inflatable structure (Hammond et al., 2017; Teng et al., 2019; Suzuki et al., 2020b), a deployable structure using coilable masts (Jensen and Pellegrino, 2001; Joosten, 2007), and a mechanical structure with reel-based actuation (Takei et al., 2011b). Pneumatic actuation was problematic for our mobile setup as it requires a tube connected to a pump or pressure tank to supply air. The deployable structure and mechanical reel-based actuation afforded much higher extension ratios, but were limited in robustness and load-bearing capabilities. We finally settled on a mechanical scissor structure due to its low-cost (compact drying rack: \$15, linear actuators: \$35 × 2) and lightweight (2 kg) components while providing sufficient structural integrity to hold the weight of a variety of common objects. In comparison with existing warehouse robots such as Kiva (Guizzo, 2008), which have a limited expandable capability as they are designed for one specific shelf, our mechanical scissor lift can move various objects by leveraging its highly expandable structure (4× expansion ratio). The actuation height (30–100 cm) was chosen to cover a wide range of standard chairs and tables, which measure 30–76 cm and 48–96 cm, respectively (Woodworking, 2019). The maximum height of the scissor structure itself can be also extended by adding more elements, such as combining two scissor structures to double the maximum height. However, such an adjustment comes with a loss in structure stability.

5.1.2 Interaction Paradigms

We deployed RoomShift in applications for supporting virtual real estate tours and collaborative architectural design, two increasingly common use cases for VR (Ibayashi et al., 2015). RoomShift supported these scenarios by enabling encountered-type haptics, whereby the robots manipulates physical objects (chairs, moveable walls, etc.) in order to adapt the physical environment to mimic the virtual user experience.

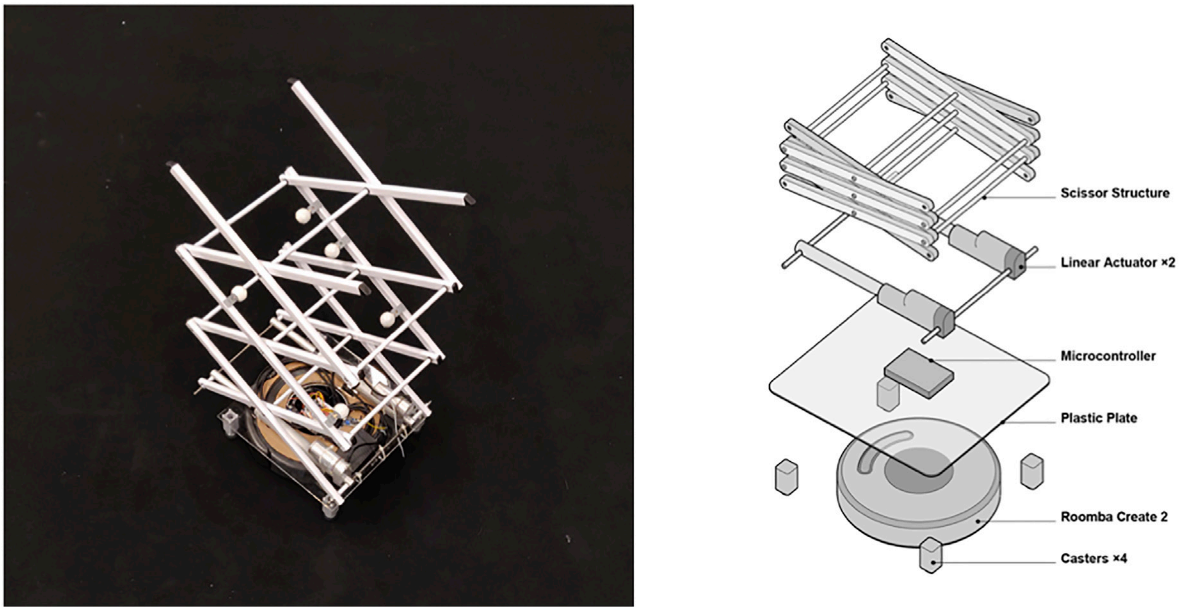


FIGURE 8 | The design of RoomShift, which integrates an expandable scissor structure with a Roomba robot.

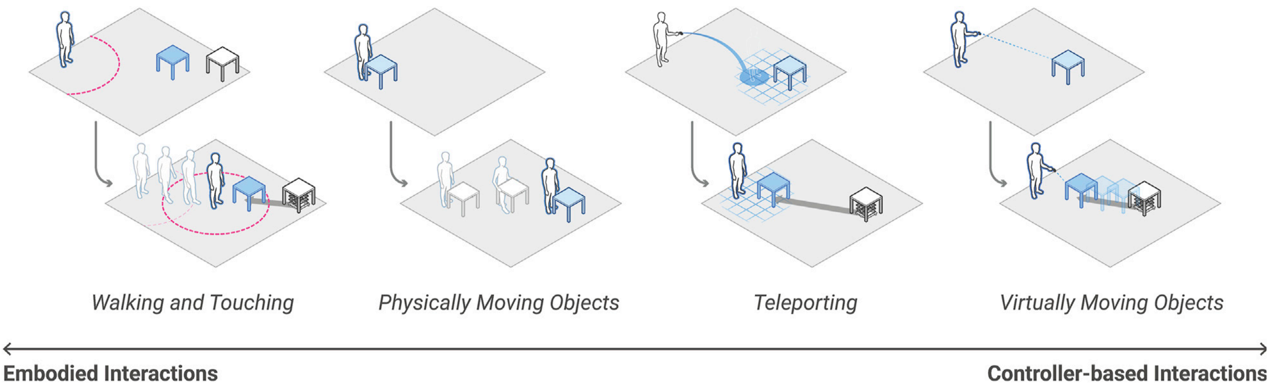


FIGURE 9 | RoomShift can provide encountered-type haptics for users in a variety of VR interactions, including when users walk to touch objects, physically move virtual objects, virtually teleport to new locations, and virtually move objects.

This system augmented several interaction paradigms for users (see **Figure 9**). For example, users could implicitly interact with the system by walking around and touching virtual objects or could explicitly interact with the virtual scene by physically moving objects tied to their virtual counterparts. Users could interact with the virtual scene with controller-based gestural interactions, for instance using controls to relocate a distant piece of furniture or removing a wall in the room. Users could also virtually teleport to new locations to navigate through virtual scenes, with RoomShift adaptively reconfiguring the physical environment to match each of the user's new virtual locations.

Traditionally, a large number of physical props and robots would be required to render virtual spaces that users can walk through and touch. Instead, RoomShift leverages low-cost, expandable structures and nine off-the-shelf robots, along with the insight that a user's immediate physical reach (e.g., ~1.5 m radius) is usually smaller than an entire virtual scene. Therefore, the system only places haptic props within the user's immediate proximity. As the user walks around the space, the robots move the props to maintain the illusion of a larger number of objects. In this way, a small number of robots with a finite set of physical props can suffice to provide haptics for the scene as the system does not need to physically render the entire environment.

In addition, the system can mimic larger objects with a single moving robot. For example, when the user is interacting with a large table, either new physical table segments can be added or a single robot can continually move the current table according to the user's position to simulate touching a larger one. This way, a limited number of robots and furniture can simulate large objects. We also use this technique for rendering larger wall segments, where the robot moves, carrying the proxy, as the user walks along the wall, similar to a technique proposed in PhyShare (He et al., 2017).

RoomShift also supports scene editing within VR. The virtual scene layout editing is similar to standard VR interactions and includes functionality like adding, removing, moving, resizing, or rotating virtual building elements and furniture with a VR controller or a GUI. For example, the user can point the controller at a virtual object and move it to a target location. RoomShift robots then update the corresponding physical object's position.

RoomShift illustrates an interesting paradigm for HRI, one in which the user does not directly interact with robots at all, but where robots seamlessly and invisibly operate in the background to augment user experiences in a manner similar to traditional goals of ubiquitous computing, where the goal of successful technologies is to fade into the background (Weiser, 1999). We conducted a preliminary evaluation of our system to gauge user responses to RoomShift (for more details on the evaluation, see (Suzuki et al., 2020a)). In a within-subjects counterbalanced design, participants interacted with physical chairs in a VR scene in two conditions: (1) with physical chairs moved by robots and (2) with static physical chairs. All participants expressed that the realism of the two conditions were the same. In general, participants indicated positive experiences and were enthusiastic about potential applications. By integrating off-the-shelf robots with inexpensive expandable structures and actuators, we added entirely new robot functionality and purpose, enabling new forms of interaction with humans working in a VR scene.

5.2 PufferBot

Next, we describe PufferBot (Hedayati et al., 2020), an example of how expandable structures can serve multiple purposes for robots, such as capturing human attention, conveying information, and mimicking human emotions, while also improving safety. PufferBot's design illustrates an integration of isokinetic structures, inspired by Hoberman spheres, with aerial robots. For PufferBot, we designed four different isokinetic expandable structures (ring, cylinder, hemisphere, and sphere) and three biologically-inspired behaviors for the structure to emulate (expansion, contraction, and pulsating). Below, we detail our PufferBot design and implementation process and summarize our findings of user perceptions of PufferBot.

5.2.1 Design and Implementation

Our goal in designing PufferBot was to explore the integration of expandable structures and aerial robots, with the notion that such structures might enable new forms of robot signaling and serve as protective guards to reduce the dangers of collisions. Previous robot design approaches have focused either on protecting (e.g., propeller guards) or signaling (e.g., alarms). Our insight was that expandable structures may offer a combination of both features.

As a result, we identified four design constraints for the expandable structures. First, they should be low weight as additional weight may reduce robot flight time or, in the worst case, render the robot unable to fly. Second, they should be easy to build. There is a limited number of primitive shapes that can easily expand without drastically changing their structure. For example, many structures utilize spheres because the shape can expand and contract with ease. Pyramids and cubes are more rare as they are complex and less conducive to shape-changing. Third, the structures should be symmetrical, both when contracted and in the expanded shape. This is because the aerial robot's flight controller is programmed with a predefined center of mass. Thus, the structures should not change the robot's center of mass in the x-y plane. Changes in the z-axis however, are easier to adjust for. Fourth, we needed to design structures such that no part of the expandable structure would ever be in the way of spinning propellers, as any interaction between the propellers and the structure would lead to robot damage and likely a loss of flight.

With these constraints in mind, we designed four isokinetic structures to surround the robot: a ring, hemisphere, sphere, and cylinder (See **Figure 11**). The ring is a Hoberman linkage that is positioned slightly above the propellers, expanding and contracting on the x-y plane. The hemisphere consists of a ring with two orthogonal half-rings positioned above it. The sphere consists of three orthogonal rings (one oriented along the x-y plane, one along the y-z plane, and one along the x-z plane). The cylinder contains the same circle as the ring, as well as a second one positioned just below the propellers.

To implement these designs, PufferBot itself is comprised of three components: an off-the-shelf aerial robot (DJI Flame Wheel F450 frame), an electromechanical actuator, and one of the four expandable structures described above (see **Figure 12**). As mentioned, one of our primary concerns in designing PufferBot was structure weight. The unmodified robot frame weight is 282 g. After mounting additional components (motors, battery, flight controller, etc.), the weight of the aerial robot accumulates to 1.2 Kg. The platform itself is capable of lifting up to 1.6 Kg of payload, meaning that the expandable structure could weigh up to 0.4 Kg.

In addition, we needed to consider how to mount expandable structures to the robot frame in a manner that did not interfere with robot mobility or other internal components. The diagonal length of the robot (motor to motor) is 45 cm. We used 4.5 inch propellers (11.43 cm), which make the total length of the aerial robot 70 cm. To attach our structures, we built a plate on top of the aerial robot that provided a surface to mount and secure an expandable structure and actuator, which can be powered by the main robot power supply (we used a 4S Lithium-ion Polymer (LiPo) battery, which gives the robot a flight time of approximately 18 min). This plate also allows us to avoid direct contact with the onboard sensors in the flight controller.

We designed a one degree-of-freedom rack and pinion mechanism capable of actuating any of our four expandable structure designs. The pinion gear located in the center rotates the four individual racks at the same time, so that the actuated racks can evenly apply the expansion or contraction force to the expandable structure in four different directions with the same

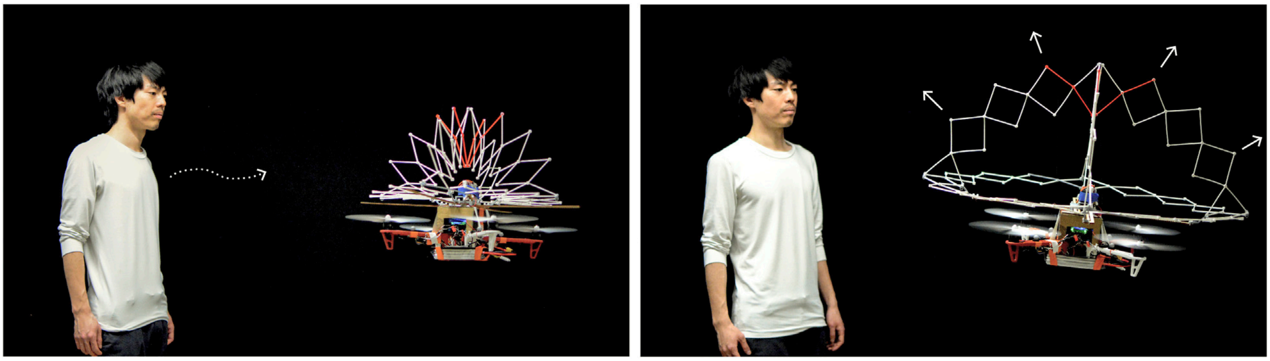


FIGURE 10 | PufferBot can exhibit various communicative behaviors when humans approach the robot. Above, PufferBot expands as a user approaches to warn the human to stay away from the robot.

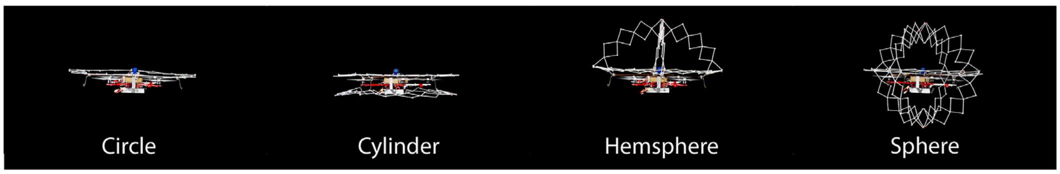


FIGURE 11 | We designed four varieties of expandable structures for PufferBot, each with trade-offs in the amount of protection it can provide, visual saliency for communication purposes, and weight.

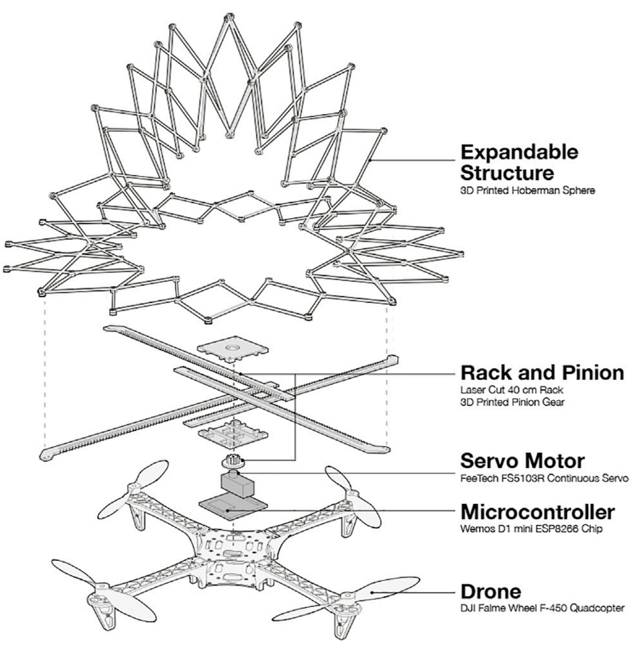


FIGURE 12 | The various components of PufferBot. At the base, we use a DJI Flame Wheel F450 drone. A microcontroller is used to control the actuation mechanism, a servo motor and a rack and pinion mechanism. The racks are joined with the expandable structure to control its radius.

magnitude. The actuator joint attached to the end of the rack can expand and collapse the expandable structure by pushing and pulling the connected points. With the mounting plate and actuation mechanism, we were able to implement each of our four structure designs in a manner that satisfied all of our design constraints. In simple flight tests, we found that each of our designs could improve human and robot safety, acting in a manner similar to deployable airbags to distribute collision forces and reduce potential propeller damage (and damage caused by propellers), often enabling the robot to remain flying even after collisions.

The robot's constraint on load capacity alongside the intended purpose of the expandable structure providing a barrier for collisions introduced a trade-off in the choice of material for the structure. While a strong, rigid structure material (e.g., metal) would provide the most protection during a collision, it would limit the allowable size of the structure, as larger structures would be too heavy for the robot to carry. On the other hand, an extremely lightweight material would be efficient in terms of load capacity, but would be more prone to break during a collision, rendering the structure ineffective. Thus, we decided to use a material that was relatively lightweight and capable of withstanding small to medium impacts and 3D printed our Hoberman linkages with PLA. We see a similar trade-off between protection and load capacity when comparing each of the four structure designs. While the sphere design offers the most protection for the robot, it is also 3 times heavier than the

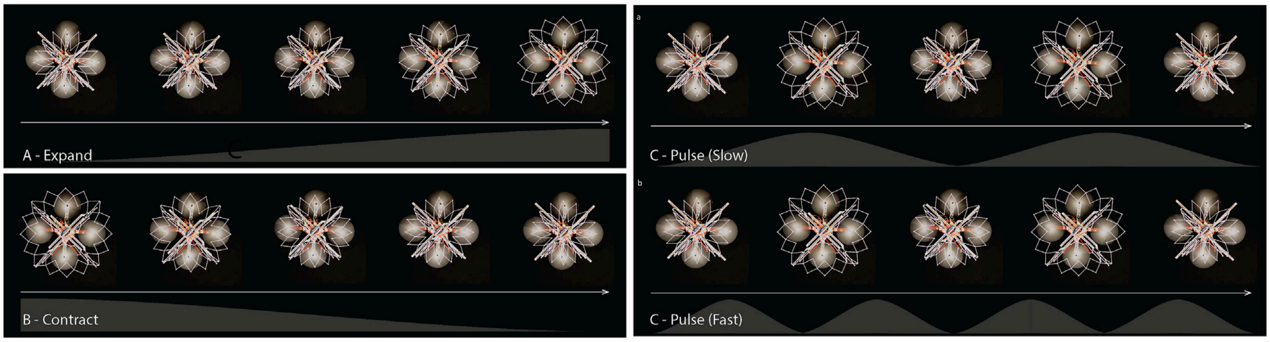


FIGURE 13 | (A) Expand, **(B)** Contract and **(C)** Pulse behaviors.

ring design. Similarly, the cylinder and hemisphere designs provide more protection than the ring, but are 2 times heavier.

5.2.2 Exploring Communicative Behaviors

Beyond safety, we were also interested in how such structures might be leveraged as a new communicative medium in human-robot interactions. To this end, we designed three expandable structure behaviors to convey information to collocated humans, with a focus on trying to convey that users should stay a safe distance away from the robot so as to avoid creating any potentially dangerous situations. In designing these behaviors, we took inspiration from nature and how animals and plants increase or decrease their size as a way to escape or frighten predators. We describe each of the three behaviors below.

Expand: This design is inspired by animals that expand in size when they want to frighten and scare off predators, like the pufferfish. Our expandable structure has a radius of 52 cm when contracted and expands to a radius of 82 cm in 6 seconds, covering the propellers (**Figure 13A**). We designed this behavior to mimic the aggressive nature of similar animal defense mechanisms, conveying a message of "don't come near me, I'm dangerous."

Contract: In contrast to the expand behavior, certain organisms contract as a defense mechanism. For example, the leaves of the shameplant (*Mimosa pudica*) fold inward and droop when touched or shaken, defending themselves from harm, and re-open a few minutes later. Such mechanisms inspired our second behavior (**Figure 13B**). During contraction, the expandable structure shrinks from an 82 cm radius to 52 cm. This behavior is intended to convey more weakness and fragility from the drone, contrasting the dangerous and intimidating nature of the expansion behavior.

Pulse: The last behavior is inspired by humans where, when a person gets anxious or afraid, their heart rate increases and they may start breathing faster. The pulse behavior consists of two sub-behaviors to mimic this physiological response. When the robot is in a room with a collocated person at a safe distance (defined as more than 3 m away (Duncan and Murphy, 2013)) the expandable structure expands and contracts at a "regular" breathing rate corresponding to approximately 20 times per minute. During this sub-behavior, the robot expands for 1 s, contracts for 1 s, and then rests for 1 s (**Figure 13C**-top). When the collocated person comes closer than 3 m to the robot, the robot starts to "breathe" faster: it expands and contracts within 1 second and takes no rest (**Figure 13C**-bottom). This is intended to indicate that the robot is anxious and the collocated person is making it uncomfortable.

We have explored people's reactions to PufferBot's expandable structure designs and behaviors by gathering information on user perceptions of various robot configurations through in-person demonstrations and online studies using recorded videos from multiple angles. We recruited 268 participants for this study: 260 for an online video study and eight for a follow-up, in-person study. In these studies, we focused on how PufferBot may dissuade users from approaching an interesting looking, but potentially dangerous robot, as well as how PufferBot can express emotions. For the online study, we asked participants to imagine that they were to approach the robot, upon which it would exhibit one of the three behaviors outlined above (i.e., expanding, contracting, or pulsing at a more rapid rate) with one of the structure designs, which they could see in a video. The participants then filled out a survey asking them to rate various qualities about the robot or their beliefs about it on a scale of 1–7. For the in-person study, participants were asked to approach the robot, which then executed one of the behaviors. In-person participants were shown all combinations of structures and behaviors and were asked to complete the same survey after each combination.

A common theme that we have found is that a majority of people (63% responses ≥5) believed the robot was discouraging them from approaching it when it exhibited any of the three behaviors. As a whole, the highest level of danger was conveyed by the expansion behavior (M = 4.42) and the hemisphere (M = 4.88). The highest level of anxiety was conveyed by the contract behavior (M = 4.92) and the hemisphere (M = 5.24). During an open-ended discussion with the in-person participants, some people believed that the robot was more dangerous in its contracted state, noting the greater exposure of the propellers. Similarly, some participants viewed the ring and cylinder structures to be unprotective of the robot or themselves due to the propeller exposure. It is important to note that even though only these two shapes were associated with a lack of protection, each shape exposes the propellers to some degree. Out of all combinations of structure shape and behavior, the in-person participants identified the sphere and heartbeat as being the most communicative or expressive. In general, these responses indicate that users may have complex and varied responses to our expandable structures and behaviors, although the structures in general may be capable of expressing simple states (e.g., users perceived the robot as experiencing noticeable levels of anxiety).

The open-ended responses from online participants also revealed complicated, and at times conflicting, user perspectives. Three participants believed that the intent of the structure was to protect the robot, rather than the human. For the ring, cylinder, and hemisphere structures, four participants thought the robot was signalling an intent to land. The responses below are illustrative of the diversity of participant opinions:

P47 (Hemisphere, Expand, Eye level) "It reminds me of a peacock expanding its feathers. It is trying to intimidate me, show me its strength. It is telling me to watch out."
P136 (Sphere, Expand, Eye level): "The robot is trying to make itself more visible so I do not accidentally crash into it."
P31 (Ring, Pulse, Below): "It almost looks like the robot is inhaling and exhaling. Like it is taking in information instead of air. As I get closer to the robot the movement seems to get faster making me believe that it is taking in more information."
P155 (Ring, Pulse, Eye level): "It seemed to be looking around for someone more interesting than me to interact with. Maybe he's trying to say, "you don't interest me."
P181 (Cylinder, Contract, eye level): "It seems to want to say 'come here with me and follow' to me."
P95 (Cylinder, Pulse, Below): "It feels like the robot is extracting something from me, and since it is not physically touching me, I feel like it is trying to extract information from my phone or personal electronic devices."

Overall, the PufferBot platform demonstrates how expandable structures and their corresponding nature-inspired behaviors might be used by robots in multiple ways simultaneously and opens the door to future research exploring the complex intersection of expandable robot structures and user responses. In the future, we hope to explore additional aspects of human-robot interaction, such as whether such structures may enhance user confidence when operating an aerial robot as crashes may cause less harm.

5.3 ShapeBots

As a final case study, we describe ShapeBots (Suzuki et al., 2019b), a swarm of small, self-transformable robots that can *individually* and *collectively* change their configurations to display information, actuate objects, act as tangible controllers, visualize data, and provide adaptive physical affordances. Each ShapeBot robot can change its individual shape by leveraging small linear actuators that are thin (2.5 cm) and highly extendable (up to 20 cm) in both horizontal and vertical directions. The modular design of each actuator enables various shapes and geometries of self-transformation. Below, we detail the design of ShapeBots, illustrate several potential application scenarios, and discuss how this type of interface opens up possibilities for the future of ubiquitous and distributed shape-changing interfaces for HRI.

5.3.1 Design and Implementation

In contrast to RoomShift and PufferBot, where our design process involved creating expandable structures and adding them to pre-existing robot platforms, we designed ShapeBots from the ground up to be robots with embedded expandable structures. Each robot is driven by two micro DC motors (TTMotor TGPP06D-136, torque: 550 g/cm, diameter: 6 mm, length: 18 mm) that are soldered to a dual motor driver (DRV8833) and controlled by a main microcontroller (ESP8266). By individually controlling rotation speed and direction, the robot moves forward and backward and turns left and right. Two 3D printed wheels (1 cm diameter) connect directly to the DC motors. An O-ring tire on each wheel increases friction with the ground to avoid slipping. A LiPo battery (3.7 V 110mAh) powers both the microcontroller and the motors.

For the expandable structure, we developed a miniature reel-based linear actuator that fits into a small footprint (3 cm × 3 cm) while able to extend up to 20 cm in both horizontal and vertical directions. The modular design of each linear actuator unit enables the construction of various shapes and geometries of individual shape transformations as seen in **Figure 14** (e.g., horizontal lines, vertical lines, curved lines, 2D area expansion with an expandable origami structure, and 3D volumetric change with a Hoberman mechanism). Such transformations support three major types of shape change (form, volume, and orientation) categorized in Rasmussen et al. (2012). Each robot has an additional DRV8833 motor driver to control these linear actuators; the two motor drivers connect to the microcontroller through a 2-sided custom PCB.

All components are enclosed within a 3D printed housing (3.6 cm × 3.6 cm × 3 cm) with three rectangular holes in the front side (**Figure 14**) that provide micro USB ports for programming, recharging, and the microcontroller reset switch. All 3D printed parts were fabricated with a FDM 3D printer (Cetus 3D MKII) and PLA filament (Polymaker PolyLite 1.75 mm True White). For horizontal extension, each linear actuator unit is fixed with a custom 3D printed holders. For the vertical extension, we used a thick double-sided tape (3M Scotch Mounting Tape 0.5 inch) on top of the swarm robot. In our current prototype, one swarm robot costs approximately 20–25 USD (microcontroller: 4 USD, motor drivers: 3.5 USD x2, DC motors: 3 USD x2, charger module: 1 USD, LiPo battery: 4 USD, PCB: 1 USD) and each linear actuator costs approximately 6-7 USD (DC motors: 3 USD x2, limit switch: 0.5 USD, polyester sheet: 0.1 USD), but this cost can be reduced with volume. For our system, we fabricated thirty linear actuator units for twelve robots. To control the swarm of robots, we implemented a custom centralized PID controller. The PID controller gets the position of ShapeBots from the unique fiducial marker attached to each of the robots using an RGB camera and sends control signals to each robot through Wifi. As an example, to create a formation (e.g., sine wave) the PID controller moves each of the robots from their current state to the desired location.

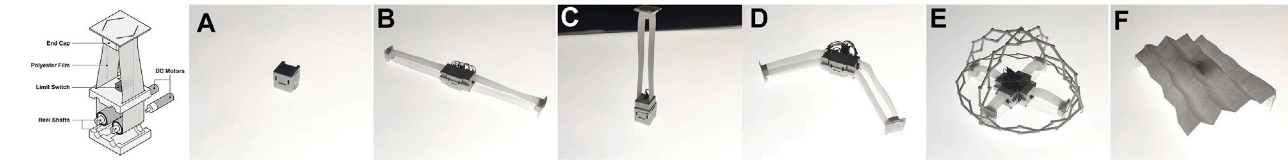


FIGURE 14 | The ShapeBot expandable structure design (left) and different types of transformation it enables: **(A)** the basic ShapeBot, **(B)** horizontal extension, **(C)** vertical extension, **(D)** bending, **(E)** volume expansion, and **(F)** area expansion.

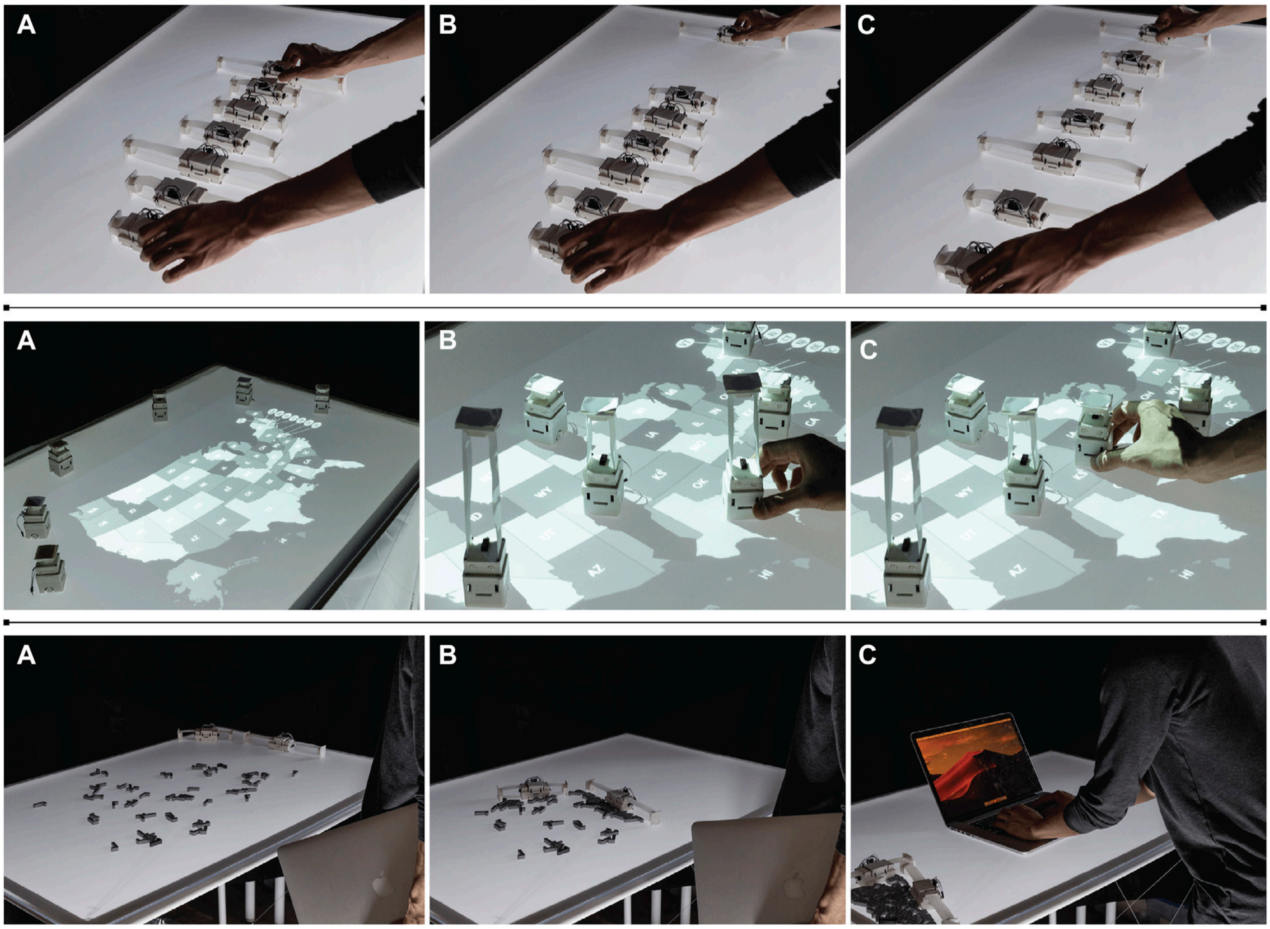


FIGURE 15 | First row: An interactive and animated sine wave. **(A)** Animated sine wave. **(B)** When the user moves one element, **(C)** then each robot can collectively move to change the spatial period of the wave. Second row: Embedded data physicalization on a map. **(A)** Projected US map. **(B)** When the user selects the dataset, the ShapeBots move to position and visualize data with their heights. **(C)** When moved, the robots change their heights accordinly. Third row: Clean up robots. **(A)** A desk is filled with debris. **(B)** Two robots starts moving and wiping the debris. **(C)** Once the robots finish cleaning up, the user can start using the workspace.

5.3.2 HRI Applications

We describe several application scenarios showing how a swarm of distributed self-transformable robots might support everyday interactions. For example, one potential application area is interactive data physicalization (Jansen et al., 2015; Taylor et al., 2015), as in the first row of **Figure 15**, where seven ShapeBots transform individually to represent a sine wave. These representations are interactive with user input: when the user moves the end robot to the right, the others move to change the wavelength. The user can dynamically change the amplitude of the wave by specifying the maximum length.

ShapeBots also support transforming data into different representations, such as bar graphs, line charts, and star graphs. Users can place and move robots, which enables embedded data representations (Willett et al., 2017). For example, ShapeBots can be placed on a map of the USA to physically represent population density by changing their height based on what state they are placed on (**Figure 15**, second row).

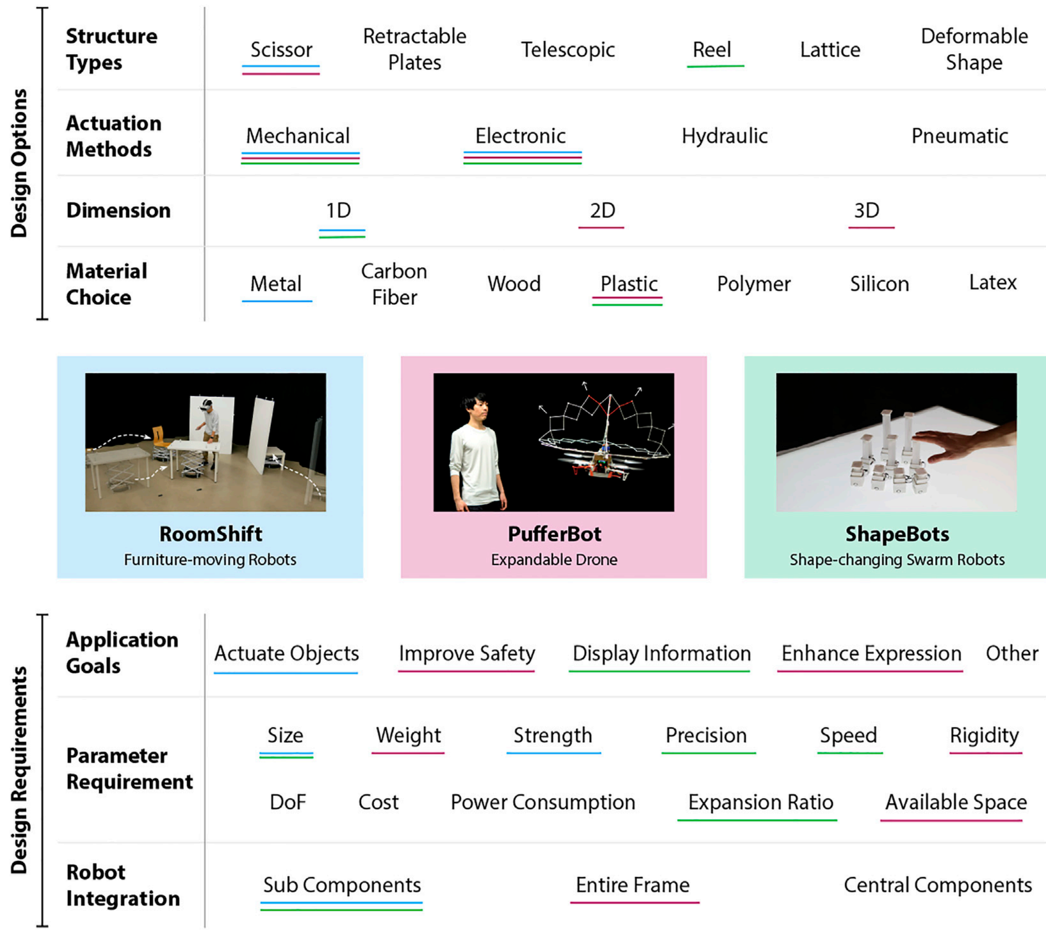


FIGURE 16 | Above, we synthesize the major implementation and interaction considerations for designing expandable structures for HRI. We also highlight the choices made in each of our three case studies.

Users can interact with the dataset by placing a new robot or moving a robot to a different state, with the robots updating their physical forms to represent the respective population.

Other examples of distributed representations include showing the magnitude and orientation of wind on a weather map or physicalizing magnetic force fields. This physical data representation might be particularly useful for people with visual impairments (Suzuki et al., 2017; Guinness et al., 2018). ShapeBots can also act as an interactive physical display, meaning they can render or enable users to preview various shapes. For instance, when reading a picture book of animals, children might visualize a fish with ShapeBots at actual size. Another application of Shapebots is for use as an interactive tangible display. As an example, four ShapeBots might display a small rectangle and, when the user moves a robot, the others can change positions and lengths to appropriately scale the shape. The user can also move robots to rotate or translate the shape. In this manner, ShapeBots might provide a physical preview of a CAD design (e.g., if a user is designing a box, ShapeBots can physicalize the actual size of the design). In such interactions, the design process and physical rendering are tightly coupled; as the user changes aspects of the design in CAD software, the ShapeBots change accordingly or the user can change the parameters of the design by directly moving robots in the physical space, and these changes are reflected in the CAD design. Finally, ShapeBots may provide practical assistance by their ability to actuate objects and act as physical constraints. As an example, **Figure 15**, third row) shows two robots extending their linear actuators to wipe debris off a table, clearing a workspace for the user.

In summary, ShapeBots are miniature tabletop robots with expandable structures that enable individual and collective shape-change. We highlight ShapeBots as an example of how robots may be designed from the beginning with expandable structures in mind and to illustrate additional collective shape-changing capabilities for human-robot interaction beyond the implicit interactions described in RoomShift.

6 DISCUSSION AND FUTURE RESEARCH DIRECTIONS

We believe that expandable structures represent a significant and underexplored avenue for HRI research. Our case studies, along with related systems such as the Triple Scissor Extender Robot Arm (Shikari and Asada, 2018) and Rotorigami (Sareh et al., 2018), demonstrate the potential of integrating of expandable structures into robotics to enrich human-robot interactions, whereby such structures may provide robots with new interactive capabilities, enable novel forms of communication, and enhance safety. In addition, we envision that such structures may serve aesthetic and experiential purposes as well, such as piquing user curiosity, increasing enjoyment, or promoting a sense of play, although we have yet to see research explore such applications of expandable structures in HRI contexts. To aid researchers and developers seeking to explore this burgeoning space, we have summarized the major design and implementation considerations for expandable structures in robotics, highlighting the choices made in our three case studies above, in **Figure 16**. We are excited to continue exploring expandable structures for HRI and further develop the initial design space outlined here as the research community begins to leverage expandable robots in new forms of interaction. Moving forward, we believe the following aspects hold particular promise for future research:

First, we believe research may more deeply explore the use of expandable structure robots in conjunction with virtual reality, as they show great value for augmenting VR experiences through encountered-type haptics. In contrast to RoomShift, where we used expandable structures to deliver haptic proxies, future work might investigate how expandable structures could act as various haptic proxies themselves. As the number of virtual objects that someone can interact with in a virtual world is essentially limitless, it is nearly impossible to design a system like RoomShift that can deliver any type of physical object to a user in a virtual environment unless the particular application is known in advance. However, expandable and/or shape-changing structures may be able to emulate a vast array of objects with which users can physically interact. Additionally, systems might afford users the capability of changing the physical shape or size of virtual objects while simultaneously feeling such transformations in their hands.

Next, we envision future work may investigate how expandable structure robots might improve users' wellness and productivity. Through our work with PufferBot, we have found that expandable structures may alter the various anthropomorphic emotions and personality traits that humans naturally ascribe to robots. We would like to explore how to leverage expandable structures to change human perceptual responses to robots in a principled manner and believe that the range of possibilities is much greater than the small subset of affective traits we have explored to date. For example, future work might examine how expandable structure robots could convey emotions such as empathy or tranquility to improve user wellness or visualize aesthetically pleasing objects, such as blooming flowers, to bring joy to users. As a practical example, expandable structures with behaviors similar to the pulse pattern exhibited by PufferBot might be used as a guide for breathing patterns, as is done in meditation practice, in a robot-guided meditation interaction. Towards improving users' productivity, we are interested in how small expandable robot like Shapebot that could integrate within user workstations might help users visually keep track of schedules, provide appointment reminders, or increase user motivation through emotive expressions.

Beyond individual interactions, we anticipate that expandable structure robots may also hold benefits for interacting with crowds. For instance, robots with expandable structures might be used to create dynamic boundaries around areas, which could change size depending on the size of the crowd. On a larger scale (e.g., crowds of thousands of people), we envision that a swarm of robots with expandable structures might be used to direct crowd movement, such as providing guidance towards exits or along evacuation routes, by expanding to block incorrect or overcrowded paths and marking available routes. In emergency scenarios, robots might also leverage expandable structures to create space for injured parties or protect privacy.

Ultimately, we envision a future where shape-changing technologies have been woven into standard robot design practices, enabling robots to dynamically adapt to users and their environment. Expandable structures can play a key role in this vision by serving as low-cost, easy-to-implement, and easy-to-control methods to augment robot capabilities. We believe the time is ripe for HRI research to examine their potential for enhancing human-robot interactions. We hope the design space and case study examples provided here will help advance and encourage further research in this area.

AUTHOR CONTRIBUTIONS

RS, HH, and DL led the implementation and data collection for RoomShift. HH, WR, and DS led the implementation and data collection for Pufferbot. RS and DL led the implementation and data collection for ShapeBots. All authors contributed to manuscript writing and revisions and approved the submitted version.

FUNDING

This work was supported by an Early Career Faculty grant from NASA's Space Technology Research Grants Program under award NNX16AR58G and by the Nakajima Foundation scholarship and the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan.

REFERENCES

Abtahi, P., and Follmer, S. (2018). "Visuo-haptic Illusions for Improving the Perceived Performance of Shape Displays," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 150. doi:10.1145/3173574.3173724

Admoni, H. (2016). Nonverbal Communication in Socially Assistive Human-Robot Interaction. *AI Matters* 2, 9–10. doi:10.1145/3008665.3008669

Alexander, J., Roudaut, A., Steimle, J., Hornbæk, K., Bruns Alonso, M., Follmer, S., et al. (2018). "Grand Challenges in Shape-Changing Interface Research," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 299. doi:10.1145/3173574.3173873

Andrist, S., Pejsa, T., Mutlu, B., and Gleicher, M. (2012). "Designing Effective Gaze Mechanisms for Virtual Agents," in Proceedings of the SIGCHI conference on Human factors in computing systems (New York, NY, USA: ACM), 705–714. doi:10.1145/2207676.2207777

Andrist, S., Tan, X. Z., Gleicher, M., and Mutlu, B. (2014). "Conversational Gaze Aversion for Humanlike Robots," in Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (New York, NY, USA: ACM), 25–32. doi:10.1145/2559636.2559666

Baraka, K., Paiva, A., and Veloso, M. (2016). "Expressive Lights for Revealing mobile Service Robot State," in Robot 2015: Second Iberian Robotics Conference (Berlin, Germany: Springer), 107–119. doi:10.1007/978-3-319-27146-0_9

Belke, C. H., and Paik, J. (2017). Mori: a Modular Origami Robot. *Ieee/asme Trans. Mechatron.* 22, 2153–2164. doi:10.1109/tmech.2017.2697310

Caluwaerts, K., Despraz, J., Işçen, A., Sabelhaus, A. P., Bruce, J., Schrauwen, B., et al. (2014). Design and Control of Compliant Tensegrity Robots through Simulation and Hardware Validation. *J. R. Soc. Interf.* 11, 20140520. doi:10.1098/rsif.2014.0520

Cao, Y., Xu, Z., Li, F., Zhong, W., Huo, K., and Ramani, K. (2019). "V. Ra: An *In-Situ* Visual Authoring System for Robot-Iot Task Planning with Augmented Reality," in Proceedings of the 2019 on Designing Interactive Systems Conference (New York, NY, USA: ACM), 1059–1070.

Castano, A., Chokkalingam, R., and Will, P. (2000). Autonomous and Self-Sufficient Conro Modules for Reconfigurable Robots. *Distributed Autonomous Robotic Syst.* 4, 155–164. doi:10.1007/978-4-431-67919-6_15

Castro, M. N., Rasmussen, J., Andersen, M. S., and Bai, S. (2019). A Compact 3-dof Shoulder Mechanism Constructed with Scissors Linkages for Exoskeleton Applications. *Mechanism Machine Theor.* 132, 264–278. doi:10.1016/j.mechmachtheory.2018.11.007

Cauchard, J. R., Tamkin, A., Wang, C. Y., Vink, L., Park, M., Fang, T., et al. (2019). "Drone. Io: A Gestural and Visual Interface for Human-Drone Interaction," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Manhattan, NY, USA: IEEE), 153–162. doi:10.1109/hri.2019.8673011

Ceccarelli, M., Cafolla, D., Russo, M., and Carbone, G. (2018). Heritagebot Platform for Service in Cultural Heritage Frames. *Int. J. Adv. Robotic Syst.* 15. doi:10.1177/1729881418790692

Cha, E., Fitter, N. T., Kim, Y., Fong, T., and Matarić, M. J. (2018). "Effects of Robot Sound on Auditory Localization in Human-Robot Collaboration," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 434–442. doi:10.1145/3171221.3171285

Cha, E., Kim, Y., Fong, T., and Mataric, M. J. (2018). A Survey of Nonverbal Signaling Methods for Non-humanoid Robots. *FNT in Robotics* 6, 211–323. doi:10.1561/2300000057

Cha, E., and Matarić, M. (2016). "Using Nonverbal Signals to Request Help during Human-Robot Collaboration," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Manhattan, NY, USA: IEEE), 5070–5076. doi:10.1109/iros.2016.7759744

Chen, S.-C., Huang, K.-J., Chen, W.-H., Shen, S.-Y., Li, C.-H., and Lin, P.-C. (2013). Quattroped: a Leg–Wheel Transformable Robot. *IEEE/ASME Trans. Mechatronics* 19, 730–742.

Coyle, S., Majidi, C., LeDuc, P., and Hsia, K. J. (2018). Bio-inspired Soft Robotics: Material Selection, Actuation, and Design. *Extreme Mech. Lett.* 22, 51–59. doi:10.1016/j.eml.2018.05.003

Das, A., and Nabi, M. (2019). "A Review on Soft Robotics: Modeling, Control and Applications in Human-Robot Interaction," in 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (Manhattan, NY, USA: IEEE), 306–311. doi:10.1109/icccis48478.2019.8974461

Del Grosso, A. E., and Basso, P. (2010). Adaptive Building Skin Structures. *Smart Mater. Struct.* 19, 124011. doi:10.1088/0964-1726/19/12/124011

Ding, X., and Xu, K. (2009). Design and Analysis of a Novel Metamorphic Wheel-Legged Rover Mechanism. *J. Cent. South Univ. (Science Technology)* 40, 91–101.

Dragan, A. D., Lee, K. C., and Srinivasa, S. S. (2013). "Legibility and Predictability of Robot Motion," in Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction (Manhattan, NY, USA: IEEE Press), 301–308. doi:10.1109/hri.2013.6483603

Duff, D., Yim, M., and Roufas, K. (2001). "Evolution of Polybot: A Modular Reconfigurable Robot," in Proc. of the Harmonic Drive Intl. Symposium, Nagano, Japan.

Duncan, B. A., and Murphy, R. R. (2013). "Comfortable Approach Distance with Small Unmanned Aerial Vehicles," in 2013 IEEE RO-MAN (Manhattan, NY, USA: IEEE), 786–792. doi:10.1109/roman.2013.6628409

Escrig, F., and Valcarcel, J. P. (1993). Geometry of Expandable Space Structures. *Int. J. Space Structures* 8, 71–84. doi:10.1177/0266351193008001-208

Falanga, D., Kleber, K., Mintchev, S., Floreano, D., and Scaramuzza, D. (2018). The Foldable Drone: A Morphing Quadrotor that Can Squeeze and Fly. *IEEE Robotics Automation Lett.* 4, 209–216.

Felton, S., Tolley, M., Demaine, E., Rus, D., and Wood, R. (2014). A Method for Building Self-Folding Machines. *Science* 345, 644–646. doi:10.1126/science.1252610

Fenci, G. E., and Currie, N. G. (2017). Deployable Structures Classification: A Review. *Int. J. Space Structures* 32, 112–130. doi:10.1177/0266351117711290

Firouzeh, A., and Paik, J. (2015). Robogami: A Fully Integrated Low-Profile Robotic Origami. *J. Mech. Robotics* 7. doi:10.1115/1.4029491

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. *Robotics autonomous Syst.* 42, 143–166. doi:10.1016/s0921-8890(02)00372-x

Guerreiro, J., Sato, D., Asakawa, S., Dong, H., Kitani, K. M., and Asakawa, C. (2019). "Cabot: Designing and Evaluating an Autonomous Navigation Robot for Blind People," in The 21st International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY, USA: ACM), 68–82. doi:10.1145/3308561.3353771

Guinness, D., Muehlbradt, A., Szafir, D., and Kane, S. K. (2019). "Robographics: Dynamic Tactile Graphics Powered by mobile Robots," in The 21st International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY, USA: ACM), 318–328.

Guinness, D., Muehlbradt, A., Szafir, D., and Kane, S. K. (2018). "The Haptic Video Player: Using mobile Robots to Create Tangible Video Annotations," in Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces (New York, NY, USA: ACM), 203–211.

Guizzo, E. (2008). Three Engineers, Hundreds of Robots, One Warehouse. *IEEE Spectr.* 45, 26–34. doi:10.1109/mspec.2008.4547508

Hammond, Z. M., Usevitch, N. S., Hawkes, E. W., and Follmer, S. (2017). "Pneumatic Reel Actuator: Design, Modeling, and Implementation," in 2017 IEEE International Conference on Robotics and Automation (ICRA) (Manhattan, NY, USA: IEEE), 626–633. doi:10.1109/icra.2017.7989078

Hardy, J., Weichel, C., Taher, F., Vidler, J., and Alexander, J. (2015). "Shapeclip: towards Rapid Prototyping with Shape-Changing Displays for Designers," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 19–28.

Hawkes, E., An, B., Benbernou, N. M., Tanaka, H., Kim, S., Demaine, E. D., et al. (2010). Programmable Matter by Folding. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12441–12445. doi:10.1073/pnas.0914069107

Hawkes, E. W., Blumenschein, L. H., Greer, J. D., and Okamura, A. M. (2017). A Soft Robot that Navigates its Environment through Growth. *Sci. Robot* 2, eaan3028. doi:10.1126/scirobotics.aan3028

He, Z., Zhu, F., and Perlin, K. (2017). "Physhare: Sharing Physical Interaction in Virtual Reality," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: ACM), 17–19.

Hedayati, H., Suzuki, R., Leithinger, D., and Szafir, D. (2020). Pufferbot: Actuated Expandable Structures for Aerial Robots. *CoRR abs/2008.07615*. doi:10.1109/iros45743.2020.9341088

Hedayati, H., Walker, M., and Szafir, D. (2018). "Improving Collocated Robot Teleoperation with Augmented Reality," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 78–86. doi:10.1145/3171221.3171251

Hu, Y., and Hoffman, G. (2019). "Using Skin Texture Change to Design Emotion Expression in Social Robots," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Manhattan, NY, USA: IEEE), 2–10. doi:10.1109/hri.2019.8673012

Ibayashi, H., Sugiura, Y., Sakamoto, D., Miyata, N., Tada, M., Okuma, T., et al. (2015). "Dollhouse Vr: a Multi-View, Multi-User Collaborative Design Workspace with Vr Technology," in SIGGRAPH Asia 2015 Emerging Technologies (New York, NY, USA: ACM), 8.

Iwata, H., Yano, H., Nakaizumi, F., and Kawamura, R. (2001). "Project Feelex: Adding Haptic Surface to Graphics," in Proceedings of the 28th annual conference on Computer graphics and interactive techniques (New York, NY, USA: ACM), 469–476.

Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., et al. (2015). "Opportunities and Challenges for Data Physicalization," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 3227–3236. doi:10.1145/2702123.2702180

Jensen, F., and Pellegrino, S. (2001). "Arm Development Review of Existing Technologies," in *Tech. Rep.* (Cambridge: Cambridge Univ.(United Kingdom). Dept. of Engineering).

Jiang, S., Patel, D. C., Kim, J., Yang, S., Mills, W. A., Zhang, Y., et al. (2020). Spatially Expandable Fiber-Based Probes as a Multifunctional Deep Brain Interface. *Nat. Commun.* 11, 6115. doi:10.1038/s41467-020-19946-9

Joosten, B. K. (2007). *Preliminary Assessment of Artificial Gravity Impacts to Deep-Space Vehicle Design.*

Kaur, M., and Kim, W. S. (2019). Toward a Smart Compliant Robotic Gripper Equipped with 3D-Designed Cellular Fingers. *Adv. Intell. Syst.* 1, 1900019. doi:10.1002/aisy.201900019

Kelmar, L., and Khosla, P. K. (1990). Automatic Generation of Forward and Inverse Kinematics for a Reconfigurable Modular Manipulator System. *J. Robotic Syst.* 7, 599–619. doi:10.1002/rob.4620070406

Kim, C.-M., and Nam, T.-J. (2015). "G-raff: An Elevating Tangible Block for Spatial Tabletop Interaction," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 4161–4164.

Kim, S., Kim, H., Lee, B., Nam, T.-J., and Lee, W. (2008). "Inflatable Mouse: Volume-Adjustable Mouse with Air-Pressure-Sensitive Input and Haptic Feedback," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 211–224.

Koditschek, D. (1984). "Natural Motion for Robot Arms," in The 23rd IEEE Conference on Decision and Control (Manhattan, NY, USA: IEEE), 733–735. doi:10.1109/cdc.1984.272106

Kornatowski, P. M., Mintchev, S., and Floreano, D. (2017). "An Origami-Inspired Cargo Drone," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Manhattan, NY, USA: IEEE), 6855–6862. doi:10.1109/iros.2017.8206607

Kotay, K., Rus, D., Vona, M., and McGray, C. (1998). "The Self-Reconfiguring Robotic Molecule," in Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No.98CH36146). vol. 1, 424–431. doi:10.1109/ROBOT.1998.676452

Kurokawa, H., Tomita, K., Kamimura, A., Kokaji, S., Hasuo, T., and Murata, S. (2008). Distributed Self-Reconfiguration of M-Tran Iii Modular Robotic System. *Int. J. Robotics Res.* 27, 373–386. doi:10.1177/0278364907085560

Kwon, M., Huang, S. H., and Dragan, A. D. (2018). "Expressing Robot Incapability," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 87–95. doi:10.1145/3171221.3171276

Lasota, P. A., Fong, T., and Shah, J. A. (2017). *A Survey of Methods for Safe Human-Robot Interaction*. Delft: Now Publishers.

Lee, D.-Y., Jung, G.-P., Sin, M.-K., Ahn, S.-H., and Cho, K.-J. (2013). "Deformable Wheel Robot Based on Origami Structure," in 2013 IEEE International Conference on Robotics and Automation (Manhattan, NY, USA: IEEE), 5612–5617. doi:10.1109/icra.2013.6631383

Li, D., Rau, P. L. P., and Li, Y. (2010). A Cross-Cultural Study: Effect of Robot Appearance and Task. *Int. J. Soc. Robotics* 2, 175–186. doi:10.1007/s12369-010-0056-9

Li, R., Liu, Z., Mi, W., Bai, S., and Zhang, J. (2019). "Design and Kinematic Analysis of a Novel Wire-Driven Spherical Scissors Mechanism," in IFToMM International Conference on Mechanisms, Transmissions and Applications, 192–200. doi:10.1007/978-981-15-0142-5_20

Liang, X., Xu, M., Xu, L., Liu, P., Ren, X., Kong, Z., et al. (2012). "The Amphihex: A Novel Amphibious Robot with Transformable Leg-Flipper Composite Propulsion Mechanism," in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (Manhattan, NY, USA: IEEE), 3667–3672. doi:10.1109/iros.2012.6386238

Liu, Y., Du, H., Liu, L., and Leng, J. (2014). Shape Memory Polymers and Their Composites in Aerospace Applications: a Review. *Smart Mater. Struct.* 23, 023001. doi:10.1088/0964-1726/23/2/023001

Ma, J., and You, Z. (2013). Energy Absorption of Thin-Walled Beams with a Pre-folded Origami Pattern. *Thin-Walled Structures* 73, 198–206. doi:10.1016/j.tws.2013.08.001

Maeder-York, P., Clites, T., Boggs, E., Neff, R., Polygerinos, P., Holland, D., et al. (2014). Biologically Inspired Soft Robot for Thumb Rehabilitation. *J. Med. Devices* 8. doi:10.1115/1.4027031

Mi, H., and Sugimoto, M. (2011). "Hats: Interact Using Height-Adjustable Tangibles in Tabletop Interfaces," in Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (New York, NY, USA: ACM), 71–74.

Miyashita, S., Guitron, S., Ludersdorfer, M., Sung, C. R., and Rus, D. (2015). "An Untethered Miniature Origami Robot that Self-Folds, Walks, Swims, and Degrades," in 2015 IEEE International Conference on Robotics and Automation (ICRA) (Manhattan, NY, USA: IEEE), 1490–1496. doi:10.1109/icra.2015.7139386

Mulu, B. (2006). "Modeling and Evaluation of Human-like Gaze Behavior," in *Humanoids' 06*.

Murata, S., Kurokawa, H., Yoshida, E., Tomita, K., and Kokaji, S. (1998). "A 3-d Self-Reconfigurable Structure," in Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No.98CH36146). vol. 1, 432–439. doi:10.1109/ROBOT.1998.677012

Nakagaki, K., Dementyev, A., Follmer, S., Paradiso, J. A., and Ishii, H. (2016). "Chainform: a Linear Integrated Modular Hardware System for Shape Changing Interfaces," in Proceedings of the 29th Annual Symposium on User Interface Software and Technology (New York, NY, USA: ACM), 87–96.

Nakagaki, K., Follmer, S., and Ishii, H. (2015). "Lineform: Actuated Curve Interfaces for Display, Interaction, and Constraint," in Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (New York, NY, USA: ACM), 333–339.

Nakagaki, K., Vink, L., Counts, J., Windham, D., Leithinger, D., Follmer, S., et al. (2016). "Materiable: Rendering Dynamic Material Properties in Response to Direct Physical Touch with Shape Changing Interfaces," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2764–2772.

Niu, S., Luo, Y., Shen, Y., and Kim, K. J. (2015). "Enabling Earthworm-like Soft Robot Development Using Bioinspired Ipmc-Scissor Lift Actuation Structures: Design, Locomotion Simulation and Experimental Validation," in 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), 499–504. doi:10.1109/ROBIO.2015.7418817

Novack, S. M., Acome, E., Keplinger, C., Alistar, M., Gross, M. D., Bruns, C., et al. (2021). "Electriflow: Soft Electrohydraulic Building Blocks for Prototyping Shape-Changing Interfaces," in Designing Interactive Systems Conference (DIS'21), 1280–1290.

Okuzaki, H., Saido, T., Suzuki, H., Hara, Y., and Yan, H. (2008). A Biomorphic Origami Actuator Fabricated by Folding a Conducting Paper. *J. Phys. Conf. Ser.* 127, 012001. doi:10.1088/1742-6596/127/1/012001

Oliveira, R., Arriaga, P., Alves-Oliveira, P., Correia, F., Petisca, S., and Paiva, A. (2018). "Friends or Foes? Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 279–288.

Onal, C. D., Tolley, M. T., Wood, R. J., and Rus, D. (2014). Origami-inspired Printed Robots. *IEEE/ASME Trans. mechatronics* 20, 2214–2221.

Paik, J. K., Hawkes, E., and Wood, R. J. (2010). A Novel Low-Profile Shape Memory alloy Torsional Actuator. *Smart Mater. Struct.* 19, 125014. doi:10.1088/0964-1726/19/12/125014

Paik, J. K., and Wood, R. J. (2012). A Bidirectional Shape Memory alloy Folding Actuator. *Smart Mater. Struct.* 21, 065013. doi:10.1088/0964-1726/21/6/065013

Park, Y.-L., Chen, B.-r., Pérez-Arancibia, N. O., Young, D., Stirling, L., Wood, R. J., et al. (2014). Design and Control of a Bio-Inspired Soft Wearable Robotic Device for Ankle-Foot Rehabilitation. *Bioinspir. Biomim.* 9, 016007. doi:10.1088/1748-3182/9/1/016007

Pellegrino, S. (2002). *Deployable Structures, Vol. 412.* Berlin, Germany: Springer Science & Business Media.

Perez-Guagnelli, E. R., Nejus, S., Yu, J., Miyashita, S., Liu, Y., and Damian, D. D. (2018). "Axially and Radially Expandable Modular Helical Soft Actuator for Robotic Implantables," in 2018 IEEE International Conference on Robotics and Automation (ICRA) (Manhattan, NY, USA: IEEE), 4297–4304. doi:10.1109/icra.2018.8461239

Pierce, B., Kuratate, T., Vogl, C., and Cheng, G. (2012). "Mask-Bot 2i": An Active Customisable Robotic Head with Interchangeable Face," in 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012) (Manhattan, NY, USA: IEEE), 520–525.

Raffle, H. S., Parkes, A. J., and Ishii, H. (2004). "Topobo: a Constructive Assembly System with Kinetic Memory," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 647–654.

Rasmussen, M. K., Pedersen, E. W., Petersen, M. G., and Hornbæk, K. (2012). "Shape-changing Interfaces: a Review of the Design Space and Open Research Questions," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 735–744.

Reid, W., Fitch, R., Göktoğan, A. H., and Sukkarieh, S. (2020). Sampling-based Hierarchical Motion Planning for a Reconfigurable Wheel-on-leg Planetary Analogue Exploration Rover. *J. Field Robotics* 37, 786–811. doi:10.1002/rob.21894

Roudaut, A., Krusteva, D., McCoy, M., Karnik, A., Ramani, K., and Subramanian, S. (2016). "Cubimorph: Designing Modular Interactive Devices," in 2016 IEEE International Conference on Robotics and Automation (ICRA) (Manhattan, NY, USA: IEEE), 3339–3345. doi:10.1109/icra.2016.7487508

Roudaut, A., Reed, R., Hao, T., and Subramanian, S. (2014). "Changibles: Analyzing and Designing Shape Changing Constructive Assembly," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 2593–2596.

Russo, M., and Ceccarelli, M. (2020). A Survey on Mechanical Solutions for Hybrid mobile Robots. *Robotics* 9, 32. doi:10.3390/robotics9020032

Sabelhaus, A. P., Bruce, J., Caluwaerts, K., Manovi, P., Firoozi, R. F., Dobi, S., et al. (2015). "System Design and Locomotion of Superball, an Untethered Tensegrity Robot," in 2015 IEEE international conference on robotics and automation (ICRA) (Manhattan, NY, USA: IEEE), 2867–2873. doi:10.1109/icra.2015.7139590

Sadigh, D., Sastry, S., Seshia, S. A., and Dragan, A. D. (2016). *Planning for Autonomous Cars that Leverage Effects on Human Actions*.

Salemi, B., Moll, M., and Shen, W.-m. (2006). "Superbot: A Deployable, Multi-Functional, and Modular Self-Reconfigurable Robotic System," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3636–3641. doi:10.1109/iros.2006.281719

Sareh, P., Chermprayong, P., Emmanuelli, M., Nadeem, H., and Kovac, M. (2018). Rotorigami: A Rotary Origami Protective System for Robotic Rotorcraft. *Sci. Robot* 3. doi:10.1126/scirobotics.aah5228

Sedal, A., Memar, A. H., Liu, T., Mengüç, Y., and Corson, N. (2020). Design of Deployable Soft Robots through Plastic Deformation of Kirigami Structures. *IEEE Robot. Autom. Lett.* 5, 2272–2279. doi:10.1109/lra.2020.2970943

Shang, H., Wei, D., Kang, R., and Chen, Y. (2018). Gait Analysis and Control of a Deployable Robot. *Mechanism machine Theor.* 120, 107–119. doi:10.1016/j.mechmachtheory.2017.09.020

Shikari, A., and Asada, H. (2018). Triple Scissor Extender Robot Arm: a Solution to the Last One Foot Problem of Manipulation. *IEEE Robot. Autom. Lett.* 3, 3975–3982. doi:10.1109/lra.2018.2859022

Shu, J., and Chirarattananon, P. (2019). A Quadrotor with an Origami-Inspired Protective Mechanism. *IEEE Robot. Autom. Lett.* 4, 3820–3827. doi:10.1109/lra.2019.2929978

Siddiqui, A., Dudnick, R., Kowalski, T., and Kowalski, T. (2007). Expandable Polyester Silicon-Covered Stent for Malignant Esophageal Structures before Neoadjuvant Chemoradiation: A Pilot Study. *Dig. Dis. Sci.* 52, 823–829. doi:10.1007/s10620-006-9513-6

Siu, A. F., Gonzalez, E. J., Yuan, S., Ginsberg, J. B., and Follmer, S. (2018). "Shapeshift: 2d Spatial Manipulation and Self-Actuation of Tabletop Shape Displays for Tangible and Haptic Interaction," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 291.

[Dataset] Snelson, K. D. (1965). Continuous Tension, Discontinuous Compression Structures. *US Patent 3,169,611*.

Song, S., and Yamada, S. (2018). "Bioluminescence-inspired Human-Robot Interaction: Designing Expressive Lights that Affect Human's Willingness to Interact with a Robot," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 224–232.

Stevenson, A., Perez, C., and Vertegaal, R. (2010). "An Inflatable Hemispherical Multi-Touch Display," in TEI '11: Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (New York, NY, USA: Association for Computing Machinery), 289–292. doi:10.1145/1935701.1935766

Støy, K. (2015). "Reconfigurable Robots," in *Springer Handbook of Computational Intelligence* (Berlin, Germany: Springer), 1407–1421.

SunSpiral, V., Gorospe, G., Bruce, J., Iscen, A., Korbel, G., Milam, S., et al. (2013). Tensegrity Based Probes for Planetary Exploration: Entry, Descent and landing (Edl) and Surface Mobility Analysis. *Int. J. Planet. Probes* 7, 13.

Suzuki, R., Hedayati, H., Zheng, C., Bohn, J. L., Szafir, D., Do, E. Y.-L., et al. (2020). "Roomshift: Room-Scale Dynamic Haptics for Vr with Furniture-Moving Swarm Robots," in CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (New York, NY, USA: Association for Computing Machinery), 1–11. doi:10.1145/3313831.3376523

Suzuki, R., Nakayama, R., Liu, D., Kakehi, Y., Gross, M. D., and Leithinger, D. (2020). "Lifttiles: Constructive Building Blocks for Prototyping Room-Scale Shape-Changing Interfaces," in Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (New York, NY, USA: ACM).

Suzuki, R., Nakayama, R., Liu, D., Kakehi, Y., Gross, M. D., and Leithinger, D. (2019). "Lifttiles: Modular and Reconfigurable Room-Scale Shape Displays through Retractable Inflatable Actuators," in Adjunct Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: ACM).

Suzuki, R., Ofek, E., Sinclair, M., Leithinger, D., and Gonzalez-Franco, M. (2021). "Hapticbots: Distributed Encountered-type Haptics for Vr with Multiple Shape-Changing mobile Robots," in The 34th Annual ACM Symposium on User Interface Software and Technology, 1269–1281. doi:10.1145/3472749.3474821

Suzuki, R., Stangl, A., Gross, M. D., and Yeh, T. (2017). "Fluxmarker: Enhancing Tactile Graphics with Dynamic Tactile Markers," in Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (New York, NY, USA: ACM), 190–199.

Suzuki, R., Zheng, Y., Clement, K., Yeh, T., Yi-Luen Do, E., Gross, M. D., et al. (2019). "Shapebots: Shape-Changing Swarm Robots," in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: ACM). doi:10.1145/3332165.3347911

Szafir, D. (2019). "Mediating Human-Robot Interactions with Virtual, Augmented, and Mixed Reality," in International Conference on Human-Computer Interaction, 124–149. doi:10.1007/978-3-030-21565-1_9

Szafir, D., Mutlu, B., and Fong, T. (2015). "Communicating Directionality in Flying Robots," in 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Manhattan, NY, USA: IEEE), 19–26. doi:10.1145/2696454.2696475

Szafir, D., Mutlu, B., and Fong, T. (2014). "Communication of Intent in Assistive Free Flyers," in 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Manhattan, NY, USA: IEEE), 358–365. doi:10.1145/2559636.2559672

Szafir, D., Mutlu, B., and Fong, T. (2017). Designing Planning and Control Interfaces to Support User Collaboration with Flying Robots. *Int. J. Robotics Res.* 36, 514–542. doi:10.1177/0278364916688256

Takei, S., Iida, M., and Naemura, T. (2011). "Kinereels: Extension Actuators for Dynamic 3d Shape," in SIGGRAPH '11: ACM SIGGRAPH 2011 Posters (New York, NY, USA: Association for Computing Machinery). doi:10.1145/2037715.2037810

Takei, S., Iida, M., and Naemura, T. (2011). "Kinereels: Extension Actuators for Dynamic 3d Shape," in ACM SIGGRAPH 2011 Posters (New York, NY, USA: ACM), 84.

Takei, S., Iida, M., and Naemura, T. (2012). "Morphys: Morphing Physical Environment Using Extension Actuators," in ACM SIGGRAPH.

Taylor, A. S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V., Grainger, L., et al. (2015). "Data-in-place: Thinking through the Relations between Data and Community," in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), 2863–2872.

Teng, S.-Y., Lin, C.-L., Chiang, C.-h., Kuo, T.-S., Chan, L., Huang, D.-Y., et al. (2019). "Tilepop: Tile-type Pop-Up Prop for Virtual Reality," in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: ACM).

Tolley, M. T., Felton, S. M., Miyashita, S., Aukes, D., Rus, D., and Wood, R. J. (2014). Self-folding Origami: Shape Memory Composites Activated by Uniform Heating. *Smart Mater. Struct.* 23, 094006. doi:10.1088/0964-1726/23/9/094006

Truby, R. L., Wehner, M., Grosskopf, A. K., Vogt, D. M., Uzel, S. G. M., Wood, R. J., et al. (2018). Soft Somatosensitive Actuators via Embedded 3d Printing. *Adv. Mater.* 30, e1706383. doi:10.1002/adma.201706383

Vander Hoff, E., Jeong, D., and Lee, K. (2014). "Origamibot-i: A Thread-Actuated Origami Robot for Manipulation and Locomotion," in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (Manhattan, NY, USA: IEEE), 1421–1426. doi:10.1109/iros.2014.6942743

Vespignani, M., Friesen, J. M., SunSpiral, V., and Bruce, J. (2018). "Design of Superball V2, a Compliant Tensegrity Robot for Absorbing Large Impacts," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (Manhattan, NY, USA: IEEE), 2865–2871. doi:10.1109/iros.2018.8594374

Vincent, J. F. V. (2000). Deployable Structures in Nature: Potential for Biomimicking. *Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci.* 214, 1–10. doi:10.1177/095440620021400101

Waldherr, S., Romero, R., and Thrun, S. (2000). A Gesture Based Interface for Human-Robot Interaction. *Autonomous Robots* 9, 151–173. doi:10.1023/A:1008918401478

Walker, M. E., Hedayati, H., and Szafir, D. (2019). "Robot Teleoperation with Augmented Reality Virtual Surrogates," in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (Manhattan, NY, USA: IEEE), 202–210. doi:10.1109/hri.2019.8673306

Walker, M., Hedayati, H., Lee, J., and Szafir, D. (2018). "Communicating Robot Motion Intent with Augmented Reality," in Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 316–324. doi:10.1145/3171221.3171253

Wang, W., Kim, N.-G., Rodrigue, H., and Ahn, S.-H. (2017). Modular Assembly of Soft Deployable Structures and Robots. *Mater. Horiz.* 4, 367–376. doi:10.1039/c6mh00550k

Wang, Z., Li, K., He, Q., and Cai, S. (2019). A Light-Powered Ultralight Tensegrity Robot with High Deformability and Load Capacity. *Adv. Mater.* 31. doi:10.1002/adma.201806849

Weiser, M. (1999). The Computer for the 21 St century. *SIGMOBILE Mob. Comput. Commun. Rev.* 3, 3–11. doi:10.1145/329124.329126

Willett, W., Jansen, Y., and Dragicevic, P. (2017). Embedded Data Representations. *IEEE Trans. Vis. Comput. Graphics* 23, 461–470. doi:10.1109/tvcg.2016.2598608

[Dataset] Woodworking, H. (2019). *Standard Furniture Dimensions.*

Wurman, P. R., D'Andrea, R., and Mountz, M. (2008). Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses. *AI Mag.* 29, 9.

Yao, L., Niiyama, R., Ou, J., Follmer, S., Della Silva, C., and Ishii, H. (2013). "PneUI," in UIST '13: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA: Association for Computing Machinery), 13–22. doi:10.1145/2501988.2502037

Yim, M. (1994). *Locomotion with a Unit-Modular Reconfigurable Robot.* Ph.D. thesis (Palo Alto, CA: Stanford University).

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto, T. (2006). "Responsive Robot Gaze to Interaction Partner," in Robotics: Science and systems (Princeton, NJ, USA: Citeseer), 37–43. doi:10.15607/rss.2006.ii.037

Zhang, M., Geng, X., Bruce, J., Caluwaerts, K., Vespignani, M., SunSpiral, V., et al. (2017). "Deep Reinforcement Learning for Tensegrity Robot Locomotion," in 2017 IEEE International Conference on Robotics and Automation (ICRA) (Manhattan, NY, USA: IEEE), 634–641. doi:10.1109/icra.2017.7989079

Zhao, J.-S., Chu, F., and Feng, Z.-J. (2009). The Mechanism Theory and Application of Deployable Structures Based on Sle. *Mechanism Machine Theor.* 44, 324–335. doi:10.1016/j.mechmachtheory.2008.03.014

Zheng, Z., Guan, Y., Su, M., Wu, P., Hu, J., Zhou, X., et al. (2013). "Development of Isomorphic Master-Slave Robots with Modular Method," in 2013 IEEE International Conference on Information and Automation (ICIA), 1290–1295. doi:10.1109/icinfa.2013.6720493

Zhou, A., Hadfield-Menell, D., Nagabandi, A., and Dragan, A. D. (2017). "Expressive Robot Motion Timing," in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA: ACM), 22–31. doi:10.1145/2909824.3020221

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/frobt.2022.720799

Check for updates

Connecting the Dots of Social Robot Design From Interviews With Robot Creators

Patrícia Alves-Oliveira[1]*, Alaina Orr[2], Elin A. Björling[2] and Maya Cakmak[1]

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, United States, [2]Department of Human-Centered Design and Engineering, University of Washington, Seattle, WA, United States

Despite promises about the near-term potential of social robots to share our daily lives, they remain unable to form autonomous, lasting, and engaging relationships with humans. Many companies are deploying social robots into the consumer and commercial market; however, both the companies and their products are relatively short lived for many reasons. For example, current social robots succeed in interacting with humans only within controlled environments, such as research labs, and for short time periods since longer interactions tend to provoke user disengagement. We interviewed 13 roboticists from robot manufacturing companies and research labs to delve deeper into the design process for social robots and unearth the many challenges robot creators face. Our research questions were: 1) What are the different design processes for creating social robots? 2) How are users involved in the design of social robots? 3) How are teams of robot creators constituted? Our qualitative investigation showed that varied design practices are applied when creating social robots but no consensus exists about an optimal or standard one. Results revealed that users have different degrees of involvement in the robot creation process, from no involvement to being a central part of robot development. Results also uncovered the need for multidisciplinary and international teams to work together to create robots. Drawing upon these insights, we identified implications for the field of Human-Robot Interaction that can shape the creation of best practices for social robot design.

Keywords: social robots, market studies, product design, qualitative research, domain experts

OPEN ACCESS

Edited by:
James Everett Young,
University of Manitoba, Canada

Reviewed by:
Denise Geiskkovitch,
University of Manitoba, Canada
Samantha Reig,
Carnegie Mellon University,
United States

***Correspondence:**
Patrícia Alves-Oliveira
patri@uw.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *05 June 2021*
Accepted: *22 April 2022*
Published: *02 June 2022*

Citation:
Alves-Oliveira P, Orr A, Björling EA and Cakmak M (2022) Connecting the Dots of Social Robot Design From Interviews With Robot Creators. Front. Robot. AI 9:720799. doi: 10.3389/frobt.2022.720799

1 INTRODUCTION

How does a designer start to create a social robot? Our work lifts the curtain on a topic thus far unexplored: how robot creators, from industry to research labs, design and fabricate social robots. We shed light on current design practices for social robots and derive specific implications for the emerging field of human-robot interaction (HRI). By identifying limitations and challenges inherent in social robot design, we intend to inspire use of best practices for their creation, helping researchers and commercial designers build higher quality products that are better suited for consumer markets. Our ultimate goal is to inspire robot creators to build social robots that are closely aligned with humans needs and values within the socio-technological society in which we live.

The term *social robot* has been used to define 'socially interactive robots' (Fong et al., 2003) that have one or more of the following competencies: the ability to communicate, express affective behaviors and/or perceive human emotions, have personality or character, model social aspects of humans, learn and/or develop social skills, and establish and maintain social relationships (Shibata, 2004; Yanco and Drury, 2004; Dautenhahn, 2007). Therefore, designing social robots requires a combined understanding and knowledge integration about human behavior and intelligence, as well as a diverse set of technical skills, e.g., in computation and fabrication (Baraka et al., 2020). This makes social robot design intrinsically interdisciplinary compared to the design of other artifacts or technologies.

In this paper, we provide recommendations for future robot creators to inspire the design of successful social robots. We conducted in-depth, qualitative interviews with expert robot creators from companies and research labs who are directly involved in the design, development, and testing of social robots. During our interviews, they disclosed their design process, the extent to which end-users were involved (if at all), and how their teams were composed.

2 RELATED WORK

2.1 Uniqueness of Social Robots

Unlike industrial robots, which have been on the market for some time, social robots are a newly emerging technology just now appearing in our stores. In this work, we deliberately chose to interview robot creators and analyze their design processes. Previous research examined how digital fabrication tools, such as 3D printers, laser cutters, and CNC routers, are fabricated by interviewing professionals that utilize these tools (Yildirim et al., 2020); it highlighted practices concerning the use of digital fabrication tools, specifically focusing on machine awareness, autonomy, and user agency. While these findings are relevant to the field of robotics—especially because the initial stages of creating a robot involve using digital fabrication tools to build prototypes—it does not explore the design process for social robots, which we address in this paper.

Sanneman et al. (2021) described the processes and challenges that companies follow when working with new technologies such as robotics and the Internet of Things (Sanneman et al., 2021). Insights from interviews with key players in the industrial robotics ecosystem contribute to research directions for the field of industrial robotics. While this work is relevant for the field of HRI since many components of social robots are shared with industrial robots—including vision, perception, and control—our work focuses more deeply on the challenges inherent to designing interactive robots that communicate with people, also fertile ground for investigation.

Previous work on robot teams explored the attitudes of frontline employees who use industrial robots every day (Sauppé and Mutlu, 2015; Elprama et al., 2017; Wurhofer et al., 2018; Welfare et al., 2019). Additionally, an extensive ethnographic investigation studied anthropomorphism in teams that work with robots (Chun and Knight, 2020). While these studies focus on the team that directly works side-by-side with robots, our work focuses on the experience of teams that design and build new robots.

By acknowledging the uniqueness of social robot design, Axelsson et al. introduced a framework for participatory design practices for social robots (Axelsson et al., 2021). This framework provides templates and guidelines to promote collaboration between multidisciplinary teams when creating social robots. This approach relates to ours; however, the authors neither explored the inclusion of users in the process of social robot design nor accounted for the benefits and shortcomings of different Human-Centered Design (HCD) practices applied to this problem, which we uncover in this paper.

2.2 Product Design and Development

The product development cycle is characterized by multiples theories and practices (Moni et al., 2020). During our research, we interviewed robot creators about the life cycle of creating a social robot. We highlight below some of the more influential practices in product design and development to better contextualize this research. Note that benefits and costs apply to all approaches, which generally work in combination rather than individually.

A *linear design process* is primarily used to manage risk when conceptualizing a product. In this practice, each phase must be fully completed before proceeding to the next, letting designers catch errors when they are least expensive and time-consuming to fix. The linear method is straightforward but requires discipline to be effective. However, during the design process, it is essential to realize that most use scenarios will require flexibility and the ability to react to new information and circumstances, challenging considerations in this linear practice (Bocken et al., 2016).

In contrast, *user-centered design (UCD)* is an iterative design process in which designers focus on users and their needs in each phase. User-Centered Design (UCD) teams involve users *via* a variety of research and design techniques to create highly useful and accessible products. There exists an explicit understanding of the users, tasks, and use environments: the aim of the process is to capture and address the whole user experience. Therefore, the design team includes professionals drawn from multiple disciplines, and experts may conduct evaluations of the produced designs using design guidelines and criteria (Still and Crane, 2017). This work uses the term *human-centered design* (HCD) to address inclusion of user emotional or psychological preferences (Gasson, 2003). Examples of HCD practices include the body of work by Don Norman (Norman, 2013) and the 7 Principles of universal Design (Story, 2001).

Finally, when using *design heuristics*, domain experts and users can assess product usability. This rapid design evaluation calls upon domain experts to go through checklists aimed at assessing the system's (or the robot's) 'heuristics' to guide future improvements (Jiménez et al., 2012). Such methods are common in the design of a variety of products and also contribute to the design of social robots.

2.3 Social Robotics Market

According to a 2020 market analysis by BCC Research, social robots are a rapidly growing market, with a compound annual growth rate of around 15% (LLC, 2020). According to Statista [1], in

[1]Statista: https://www.statista.com/statistics/755677/social-and-entertainment-robot-sales-worldwide/.

2018 social and entertainment robot sales reached 2.68 million units worldwide. By 2025, that number is forecast to double to 5.51 million units. Thus, we see an emerging recognition of the potential of this field, especially for use in healthcare, education, and entertainment.

Despite this growth, many robot companies have failed in the market. In 2019, the *Robot Report* [2] released information about social robot companies that ceased production; it noted that many initially thriving companies failed to succeed over the longer run. Our work aims to be a conversation-starter on the topic of social robot failure versus success by lifting the curtain of a topic that is often discussed but little studied in the HRI community: the inner workings of social robot design.

3 METHODS

After having developed our own social robot prototype (Björling and Rose, 2019) for the Ecological Momentary Assessment Robot (EMAR) project, we wanted to learn from other researchers and designers of scalable robots currently on the market to assess how best to scale our prototype. Our research questions were:

- What are the different design processes for creating a social robots?
- Are users involved in the design of social robots, and how?
- Who participates on the teams that develop social robots?

3.1 Collective Case Study

We conducted an *interview-based qualitative case study*, an established research design method, where we consulted a variety of stakeholders involved in the creation of social robots. Case studies enable in-depth appreciation and multi-faced exploration of complex issues or phenomena of interest (Crowe et al., 2011); their value in research lies in their ability to explain, describe, or explore events (Yin, 2017). Unlike experimental designs, which focus on testing a specific hypothesis by deliberately manipulating interventions or conditions, the case study approach captures information of a more explanatory nature by focusing on 'how,' 'what,' and 'why' questions.

Our case study focused on identifying and comprehensively describing how social robots are created. Specifically, we applied a *collective case study approach*, which simultaneously explores multiple cases in an attempt to generate broader appreciation of particular issues (Stake, 1995). Thus, our collective case study included employees from multiple social robot companies as well as professors from university labs that build social robots.

3.2 Ethics and Permissions

This study was reviewed by and received Institutional Review Board approval from the University of Washington, Seattle, WA, United States. Participants verbally consented to be recorded. To protect participants, all information that could potentially lead to identification of individuals was removed, and transcripts were anonymized.

3.3 Sampling and Recruitment

We used the purposive sampling technique to recruit subjects for this study. *Purposive sampling* is a form of non-probability sampling in which researchers deliberately choose participants due to their unique qualities (Tongco, 2007); it is one of the most effective techniques to study a specific domain with knowledgeable experts, which is the case for our study. Participants were identified by 1) drawing on the extended network of authors of this paper, and 2) applying the inclusion criteria that recruits required hands-on experience in creating social robots. We specifically focused on subjects who worked with social, interactive robots rather than industrial ones: these two markets require different knowledge and experience, are associated with different application scenarios, target different users/consumers, and exhibit diverse market maturity, with industrial robots being used in the marketplace for far longer.

We identified 18 subjects who had created one of more robots in the context of a research lab or in industry. We sampled purposefully for maximum variability, ensuring representation from a range of countries, professional backgrounds (including engineers, designers, artists, system developers, academics, and visionaries/futurists), and types of robots (Patton, 2005). The initial recruitment email included an invitation to participate in an interview about their role in the design of a social robot along with sample questions from the interview template. A total of 13 subjects expressed an interest; the follow-up email contained the interview schedule; refer to the demographic description in **Table 1** and the robots build by these creators in **Figure 1**. The remaining 5 subjects who did not enroll in this study mentioned either their concern about discussing the topic given their non-disclosure agreement (NDAs) (2 subjects) or simply failed to reply to our email solicitation (3 subjects). Industry representatives were particularly uncomfortable with sharing potentially sensitive commercial information. We stopped recruiting additional participants when we reached thematic saturation, which occurs when no new themes emerge during analysis (Guest et al., 2006).

3.4 Data Collection

Among the wide range of qualitative methods, we specifically chose to conduct interviews, which enable flexibility during data collection while remaining grounded in a particular framework (Gill et al., 2008). Interviews were conducted over Zoom [3], digitally recorded, and transcribed. They ranged from 30, –,90 min depending on the subject's availability and how in-depth the interview went. We explored the most promising areas, including the development process of robots, user or customer involvement in the robot creation, and team

[2]Robot Report 'Remembering robotics companies we lost in 2019': https://www.therobotreport.com/robotics-companies-we-lost-2019/.

[3]Zoom: https://zoom.us.

TABLE 1 | Demographic description of study subjects. All subjects are robot creators who designed and built one or more robots in a research lab or company.

Subject ID	Gender	Country	Role
1	Male	Poland	Co-Founder and Electrical Engineer
2	Male	Israel	Industrial Designer & Hardware Engineer
3	Female	Australia	Designer and Animator
4	Male	US	Robot Animator
5	Male	US	Software Engineer
6	Male	US	Software Engineer
7	Female	US	Chief Operating Officer (COO)
8	Female	US	Software Engineer
9	Male	United Kingdom	Co-Founder and Designer
10	Male	US	Senior Engineer
11	Male	US	Founder and Principal Investigator
12	Male	US	Researcher
13	Female	Portugal and US	Researcher

composition of the robot builders. A sample interview guide is shown in the box below. While several pre-defined questions employed a blend of closed- and open-ended formats, we often accompanied these with follow-up 'why' or 'how' questions. This enabled flexibility to explore novel topics raised by the subjects while having a template that guided the discussion (Dearnley, 2005; Newcomer et al., 2015).

Interview Sample. Questions were distributed according to the themes of interest in this study.

- *What was the workflow that you followed for the development of the robot?* (robot development process)
- *What types and how many robot prototypes did you explore?* (robot development process)
- *How different were the prototypes compared to the last version of the robot?* (robot development process)
- *Were users or consumers involved at any stage of the robot creation?* (user involvement)
- *What type of data was collected, and how did it inform the development of the robot?* (user involvement)
- *What were the backgrounds of the team that created the robot?* (team composition)
- *How was the division of labor distributed across the team?* (team composition)
- *With this particular robot today, what are the current pain points?* (lessons learned)
- *What would you do differently if you were to do this again from the beginning?* (lessons learned)

3.5 Data Analysis

We identified different design processes for robot creation, several degrees of user/customer involvement in the creation, and different approaches to team composition of robot builders. We anchored our data analysis in qualitative research methods, suitable methods for exploratory studies such as ours that support inductive practices; these methods can lead to prominent emerging themes without existing prior hypotheses (Sofaer, 1999). While quantitative research could potentially be useful, a growing consensus indicates that they are ideal to justify research findings or differences across samples (Park and Park, 2016). In contrast, qualitative research is concerned with aspects of reality that cannot be easily quantified, focusing on the understanding and explanation of a phenomenon and thus deepening our comprehension (Queirós et al., 2017). This was compatible with the goal of this study, which aimed to deepen the understanding of the processes and approaches used to design social robots.

Transcribed interviews were uploaded to Miro [4], an online collaborative whiteboard suitable for research analysis that enables visual organization of data and exploration of prominent themes. Three researchers were involved in collaborative coding of the data. Two researchers independently organized the interview materials into emerging themes. To ensure consistency across coders, calibration exercises were performed until stability was reached (Krippendorff, 2009). After coding 30% of the data, the two coders met to resolve discrepancies (Campbell et al., 2013); they compared their coding schemes to ascertain concordances (i.e., alignment in definitions, language, and coding logic). When discrepancies arose, a 'negotiation agreement' was used, whereby they verbally discussed differences with a mutual effort to reconcile disagreements and divergence (Hoyle et al., 2002; Garrison et al., 2006). The third coder joined the discussion when 50 and 100% of the data was coded to help disambiguate negotiations.

We approached the analysis with an initial coding framework based on our research questions to provide an initial structure to our findings. We used an *affinity diagram* approach to code and organize the data (originally called the KJ method) (Kawakita, 1991). Affinity diagramming is a technique used to externalize, make sense of, and organize large amounts of unstructured, far-ranging, and possibly dissimilar qualitative data (Hartson and Pyla, 2012). Data collected in our interviews occasionally ventured in directions that differed from our primary research focus due to the nature of open-ended questions and semi-structured interviews. While these extraneous data were interesting, if

[4]Miro: http: miro.com.

it was not relevant to our research questions, it was not included in our results.

4 RESULTS

We now present our findings on several robot development processes, the different degrees of user/customer involvement, and the team composition of robot builders.

4.1 Robot Design Process

Three categories of design processes can be derived from our data: 1) iterative, 2) linear, 3) and data-point-driven (see **Figure 2**). We extracted the most salient details of each development process to provide a deeper understanding of the various workflows. Further, these categories are not mutually exclusive, and certain aspects overlap with others, meaning that the same robot can be mapped to more than one design process. We do not aim to compare the effectiveness of one design process to another or to state preferences; rather, we provide a comprehensive illustration of the current way robots are being designed, which has inherent value.

4.1.1 Concrete and Linear Stages of Social Robot Design

A *linear robot design process* refers to concrete and sequential stages of a workflow underlying the robot's creation. Drawing from key tenets of linear processes in design, each phase in the development life cycle should be completed before moving to the next. An example was described by Subject 10 in five well-defined stages: 1) hardware exploration, which consists of creating initial prototypes and sketches of physical features for the robot, 2) design investigations, which involves experimenting with simple robot behaviors within interaction scenarios, 3) expressivity

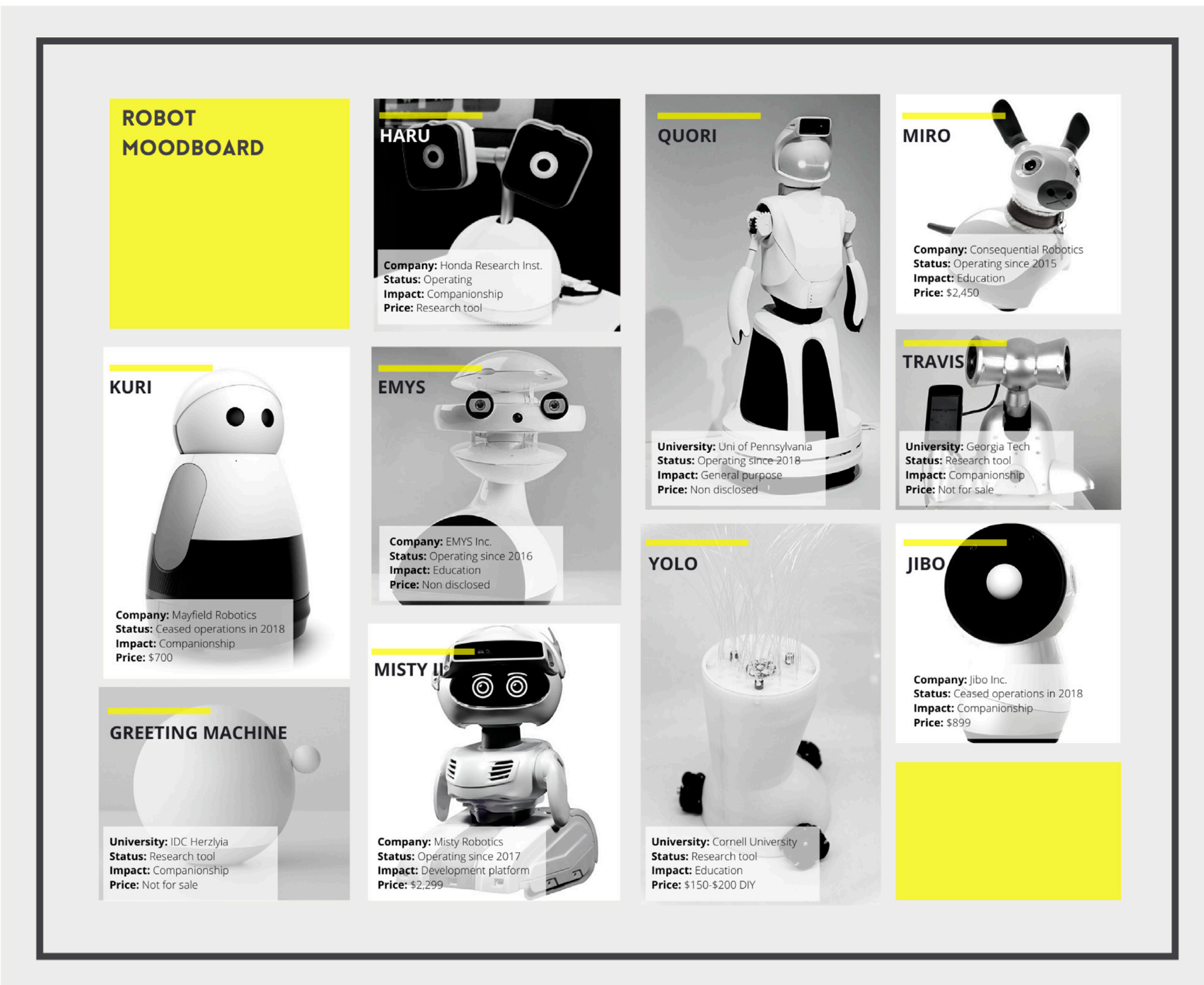


FIGURE 1 | Moodboard of robots included in this study, including details of the robot's founding company or university, business status, area of impact, and selling price. Prices may vary from those noted here.

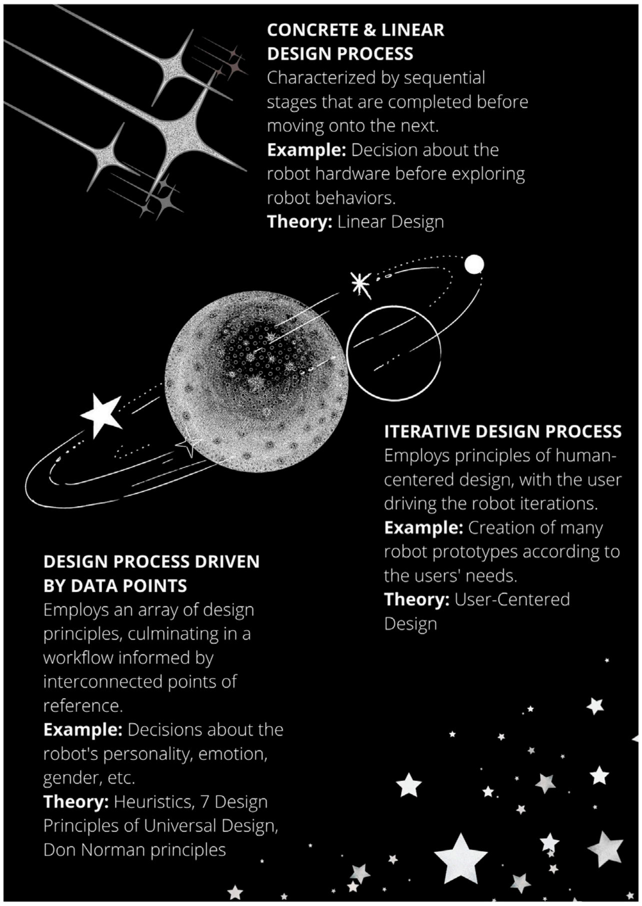


FIGURE 2 | Design process of social robots.

implementation, which consists of creating 3D printed mock-ups to test and refine degrees of freedom, 4) interaction design, which uses a puppeteer and stop motion artists to test which degrees of freedom are actually needed in the robot and which can be removed, and 5) negotiation, where the team navigates conflicting aspects of the design, striving to balance market viability and mechanical feasibility. A description of five of the created robot prototypes follows:

> *"Our first [prototype] was just a base platform. The question was: can we make it work? Our second [prototype] was this concept we liked, but we couldn't actually get it to be expressive enough. The third one has the degrees of freedom in a different place. The fourth one was pretty much the robot that you probably see now. The fifth one has some tweaks here and there, and, beyond that, there were other little tweaks."* (Subject 10, Male)

These sequential stages were described as one leading to the next, with the ultimate design decisions driven by *"this tension between how much expressibility we want the robot to have, how much it is going to cost us, how much can we sell it for, and how do we want people to interact with it."* (Subject 10, Male).

The development process described by Subject 6 included not only an iterative process but also linear stages. The overall process was conceptualized as a series of phases, *"We did go through several phases of the robot [development]. Each one of those phases lasted a few months, maybe* 4 – 5 *months each"* (Subject 6, Male). The first stage was described as *"bare bones utilitarian"*, which consisted of a prototype with finished electronic components and an unfinished exterior. This featureless version was used to evaluate the electronic components, tasks, and flows of the robot. The following three versions were assigned labels by the manufacturing team and used to perform lifecycle and long-term testing and to make various other refinements. Each phase of building was completed before moving on to the next.

Another example of a design process with clear sequential phases consists of phases that can be charted along a timeline of 4 years:

> *"In year 1, lots of prototyping, need-finding, taking prototypes out to get customer feedback, working with the industrial designers, working with mechanical designers, prototyping navigation software and drive trains and animation for the degrees of freedom and depth sensors. In the second year we hired our VPs [vice-presidents], developed alpha versions of the robot, depth sensor, at the end of the year prepped to launch at CES[5] with our painted prototypes. In the third year we launched at CES, did design for manufacturing, and at the very end of the year, shipped our first small batch of robots out to pre-order customers. In year 4 we scaled up production, got shut down just as we had our full-speed manufacturing line set up to turn out thousands of Kuris every month."* (Subjects 7 and 8, both Female)

4.1.2 Iterative Process of Social Robots Design

The*iterative development process* enables continuous improvements of the robotic system and a deeper understanding of users and their needs. A subject described this iterative design process in the following way:

> *"We went through a lot of iterations of sketching and then some low fidelity prototyping with cardboard. And the process used a lot of increasingly high fidelity prototypes with constant feedback, preferably from users. It's a kind of classic user-centric design process in many ways."* (Subject 2, Male)

Another subject described a similar iterative process. The process began with aiming for simplicity and speed because *"the first prototype is going to suck anyway and you will miss*

[5]Consumer Electronic Show (CES) is an annual trade show organized by the Consumer Technology Association (CTA) that typically hosts presentations of new products and technologies in the consumer electronics industry. Link: http://ces.tech.

the target" (Subject 1, Male). In this case, each iteration brought with it more learning, failure was acknowledged to be part of the development cycle, and the finished robot design resulted from numerous prototypes:

> *"[the development process of the robot] was very, very, very iterative. That's one of the things that, you know, when you start building a robot, you think you will build one, and NO. You will build many, many, and many more than what you think you will."* (Subject 2, Male).

He added:

> *"You do the first [prototype] and it doesn't work or works badly. Then, you learn and make it again, and it gets better. By the time you make it ten times, it's pretty good."* (Subject 2, Male)

The underlying approach of the iterative development reflects the idea that iteration strengthens one's knowledge and culminates in *"a good enough robot that works"* (Subject 2, Male). Furthermore, this approach addresses user needs and identifies pain points: *"You will want to build something that is meaningful, and that means building something that people actually want and that can solve a problem in the world."* (Subject 2, Male).

The users of social robots, though central, are not the only references that developers rely on. For example, internal team feedback was considered crucial since the team is also composed of expert roboticists that can contribute to the problem. The team would build mini prototypes of a certain robot feature or interaction case, and the internal design teams would review them and offer suggestions. They would then incorporate feedback into the next prototype and repeat the process, as noted by Subject 6:

> *"And we'd iterate back and forth on the interaction design of the skill, and then we'd go off and sort of build another version and show it to them [the team] on an average of every 3 weeks or every month kind of cadence."* (Subject 6, Male).

The iterations were driven by the goal of building a robot with minimal areas of weakness: *"at least five iterations [of the robot], and then one big robustness redesign where we're trying to fix all the things that just don't work well"* (Subject 12, Male). The iterations can be driven by a specific feature to improve in the robot. For instance, while a robot can have multiple iterations, the main motivation for them was to test different designs and physical materials, not to make users the focal point of the orbit: *"We have possibly 10 versions of the robot before we have the one that we have now and, you know, we have 10 iterations of those, like different mechanical designs, different materials, just small things."* (Subject 12, Male).

Subjects 1, 2, 6, and 12 took iterative approaches when building robots; however, Subjects 6 and 12 described workflows different from those of Subjects 1 and 2 since their orbits were more focused on gathering feedback from internal teams or improving specific features, rather than being driven by user feedback. Thus, an iterative design process does not necessarily mean it is user-centric, but rather that different expertise can be considered in the iterative process (such as the internal expertise of the team).

4.1.3 Data-Point Driven Robot Design

In a data-point-driven robot design process, the workflow is informed by points of reference, such as prior knowledge or accumulated observations. Instead of concrete stages, data points do not adhere to a specific timeline; rather, they serve as important references throughout a whole development process. Various theories and principles are associated with this method, such as Don Norman's definition of affordances, the Seven Principles of universal Design, and Jakob Nielsen's Ten Usability Heuristics, showing how this design process pulls data points from different disciplines, frameworks, and sources of knowledge.

An example was described by Subject 3, who drew on various data points when designing the robot, including her own background in design and animation, affinity for understanding the human experience, and prior research in the field of robotics. For Subject 3, the philosophy around developing social robots is that they should spark feelings and emotions in the people who use them. She mentions the importance of having companies that support this approach: *"They're brave to design a robot that brings joy"* (Subject 3, Female). Along with this philosophy was the importance of designing communication for joyful interactions mapped onto the robots' specific features and components. Subject 3 highlighted the importance of communication beyond speaking, *"What do we do when we're being social? We don't just talk. What's the content of that talk, and what are the things we do?"* (Subject 3, Female). Elaborating on this idea, creating a social robot that brings joy is exemplified in her knowledge of human behavior, which served as data points that drove the aesthetic and gesture selection design of the robot:

> *"Many of the robots in the marketplace are designed for movement, which is just motion. And not designed for gesture, which is emotion."* (Subject 3, Female)

Subject 3 explained how prior experience informed her approach to aesthetic robot design as follows:

> *"I come from an animation and design background. In my design, I always worked to aesthetic principles, which are super important to robotics. Aesthetics are important to understanding a robot's purpose, what it should be used for, and how somebody relates to a robot."* (Subject 3, Female).

An additional reference point elaborated upon by Subject 3 was the key insights from surveys and interviews with the target audience that would use the robot being created. A major finding dealt with people's expectations and comfort with social robots:

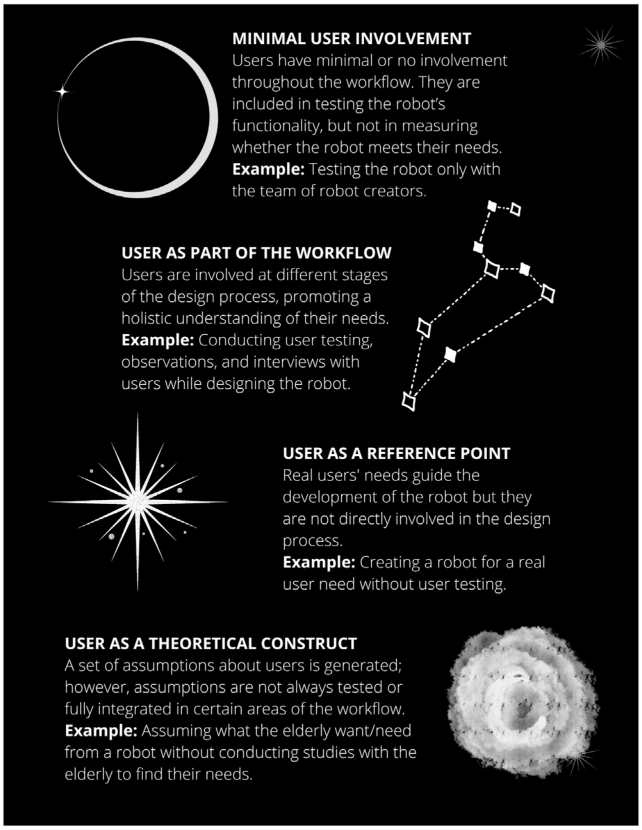


FIGURE 3 | Different degrees of user or customer involvement in the development of a social robots.

> *"In surveys I've done, there's been a rebellion against perfection in a robot, and rebellion against the robot being a know-all, and also a rebellion against humanoid robots. Also, that they [the target audience] didn't want robots to be gendered."* (Subject 3, Female)

Overall, these data points 'interacted' with one another to form the larger approach of developing a robot. Subject 3's philosophy, background, and research insights were not disparate elements of robot creation; they were instead interwoven data points that the team drew upon when designing the robot.

Another example came from Subject 9. For instance, research informed the placement and design of the robot's eyes:

> *"Stereo vision is very important. A rabbit has eyes on the side for predators. The same for cows and horses. Instead, cats and lions look straight ahead. So, the peripheral placement is less threatening and is cuter. This part is informed by research, and the robot was made with peripheral eyes. However, we placed the cameras to see straight ahead and not lose robot's functionality."* (Subject 9, Male)

Insights derived from researching animal features revealed the importance behind eye placement. Since this robot was intended to be *"a cute animal"*, mainly to have impact on the educational sector, its eyes were positioned on the side of the face to be perceived as approachable and friendly. Subject 9 also relied on his basic understanding of human behavior as a point of reference in the design process. The decision to model the robot after a pet was driven by his perspective that humans view their pets as companions:

> *"Everyone talks to their pets. Find me someone that does not talk to a pet. But does the pet understand?"* (Subject 9, Male)

An important data point is the background and experiences of robot builders. Arguably, each subject drew upon unique experiences and knowledge when designing the robots. Subjects 10 and 11 explicitly mentioned expertise as a driving force. For example, *"when the focus was hardware, the improvements were made based on failures and expertise"* (Subject 11, Male). Similarly, relying on their background in robotics and HRI when faced with design decisions was important: *"A lot of [the development process of a robot] was all about just trying to take the lessons learned. I've been in HRI for how many years, and it is like, hey, this is what people have shown so far. How do we apply these principles?"* (Subject 10, Male). Thus, a main data point is the team's expertise and unique background, which can inform the process of robot development and the underlying decisions made.

4.2 User Involvement in Robot Design

We identified three degrees of user or customer involvement in social robot development (see **Figure 3**). We note that in some stages of the robot's development, users can be involved in more than one way. Further, it is not our intention to map specific robots to specific degrees of user involvement.

4.2.1 Minimal User Involvement

This category includes robot development workflows that consider minimal input from users during robot design and testing. Here, users lack a concrete presence/identity and are brought in to test the robot's functionality, not to measure whether the robot meets deployment needs (e.g., cost, engagement in the interaction). In this case, users' involvement is limited or nonexistent throughout robot development. Subject 10 mentioned that the team did not have a specific target user in mind throughout design and testing. This can be seen in the minimal user involvement, especially in the testing and evaluation stage:

> *"Mostly, what we were able to do was show to other people in the company who weren't working directly on this robot. . .Sort of like grabbing another person in your lab and being like, 'Hey, look at this.' It wasn't as formal as a user study."* (Subject 10, Male)

In this case, the robot was tested within the team instead of with actual users or customers who would be buying and using it. In terms of design, most of the larger decisions were made by the team without direct user input. For instance, the decision to include prominent feminine features in the robot was made solely by the team, none of whom were women. This decision was a topic of contention among customers and stakeholders:

> *"We got push-back because the silhouette [of the robot] has a very thin waist and a prominent chest area. People sort of mentioned that was perhaps feeding the stereotype. There's this whole thing about gendered robots and gender perceptions, and we probably could've done a better job with that aspect."* (Subject 10, Male)

As such, not involving the users or customers who represent specific demographics in the design stage negatively affected the outcome and resulted in perceptions of stereotypes perpetuation that were neither initially intended nor considered. This was identified as a pain point and a reflection for different future decisions when creating robots: *"I would probably fight harder to bring other voices to the table. I think there's a strong view of 'Oh, we know what we want to build,' and less input from potential customers."* (Subject 10, Male). This demonstrates the importance of defining the target user to guide the direction of robot design, gathering insights on their needs, and including them in design evaluation to ensure their needs are met.

Another example of a low-level of user involvement in development is stated by Subject 6: *"I don't think it was a super user-driven design. We didn't have a ton of users."* (Subject 6, Male). Instead, the design was driven by *"the simplest possible mechanism that could still give us a wide range of expression and expressive motion"* (Subject 6, Male). Here, the workflow behind this robot seemed to prioritize optimization over user involvement.

4.2.2 User as Part of the Workflow

Through the lenses of HCD, users are central parts of the design flow when they are included in different parts of the robot development process. Knowledge gathered from the users at different design stages enables a holistic understanding of their needs, much like an outline or pattern of what would be desired in a robot. This informs the research with a specific population of user views to reflect in the design requirements (Cooley, 1999; Buchanan, 2001). For instance, Subject 3 uncovered many key insights from user input when designing the robot, which helped her understand what users want in a robot and why:

> *"I did a survey on the robot with the target age group. And what was interesting about that age group is that they didn't want a robot to share. They wanted a robot for themselves. They wanted that robot in their room. They wanted it because they spent a lot of time in their rooms. They wanted a study buddy but also wanted something that I suppose wasn't threatening. It was like social media and everything, but something that was kind of like their own friend, that was just theirs."* (Subject 3, Female)

User insight drove design decisions concerning how to convey the robot's purpose, character, and story. Instead of asking customers for their desired features, the team let users drive the design by investigating the underlying feelings behind companionship and how they might interact with a robot. From there, the team created physical representations of the user's feelings. For instance, the need for a non-threatening robot that could act as a study buddy or friend informed the design of communicative features, non-humanoid design, and genderless identity.

Subject 2 took a similar approach in centering the user in the workflow. One process involved designing a companion robot for the elderly population to help them cope with loneliness, as expressed below:

> *"We did a lot of interviews with the elderly where we showed them different types of robots, and we did thematic coding of what they think of these different robots. We then made guidelines for designing a complex social robot for the elderly, and a few things came up in the end that we used as guidelines for the design of other robots."* (Subject 2, Male)

Through interviews with the target demographic, they were able to drive the design of the robot through guiding design requirements. From the user's input, they *"designed a social robot, which has social features aiming at giving people the feeling of being seen."* (Subject 2, Male), which was one of the most salient needs amongst the elderly population they interviewed. Furthermore, the users were continually involved throughout the other stages of design. In addition to using interviews, robot developers applied other methods to understand the users and gather their input. For instance, they *"took videos of different robots that represent different kinds of robots. We showed them a few different robots and got information from them."* (Subject 2, Male).

Different methods were used according to the competencies of different users: *"For 2 years, we observed how children play with the robot prototypes that we gave them. They were always part of the process; it just didn't work asking them what the liked or not, we just need to sit still and observe. This would tell us what needed improvement."* (Subject 13, Female). In this way, the user was a central part of the workflow and drove robot design. The different inputs gathered at different design stages enabled a constellation of knowledge about the users' needs, desires, and wishes for the robot.

4.2.3 User as a Reference Point

Users can also act as reference points that drive the development of a robot. In this case, the design process consists of a fine balance between user input, designer's decisions, and business or time constraints. For instance, Subject 1 elaborated on the intricate dance of including customers in the development of the product while maintaining the designer's vision: *"The idea that you have is still important, and it is a very, very fine balance because you have to make sure you are building something for the users, but* *it is also very easy to ask the wrong questions here."* (Subject 1, Male).

After asking users what they want, robot builders can be inundated with complex features ideas that increase robot costs. There might be reluctance to let users drive the design process since the process can become subject to the 'feature creep' phenomenon [(Thompson and Norton, 2011)]:

> *There is a name in the start-up world, which is 'feature creep,' where basically you keep on adding, and adding, and adding, and adding features to your product because, well, your customers are asking. It's actually a huge risk, and it kills companies because it takes time to develop features, and it's very, very costly to develop something that's wrong because it's something that the users say they want or need, but it's not something that they would need so much that they would pay for your product."* (Subject 1, Male)

Feature creep introduces the dilemma that exists when asking for user input. On one hand, it is important to ask for and integrate user feedback in order to meet their needs. On the other hand, including user input runs the risk of adding features to the point of driving up costs, developing something that is not marketable, and creating an overly complex product. Instead, there are ongoing negotiations about finding balance between user involvement and the designer's visions for the robot. One way to achieve this balance can be to include users in overcoming major pain points of the robot's design:

> *"The way to think about that is to not build something that does everything, but to have the design process set up so that you actually build for one thing that is very, very, very specific and that solves a very big pain of your customers or users."* (Subject 1, Male)

They refer to this approach as *"solution viability"*, which is related to the idea of *"building something that is so good and solves such a big problem that people are actually willing to give you money for it, regardless of whether you ask them for this money or not."* (Subject 1, Male).

4.2.4 User as a Theoretical Construct

The user can also be included as a theoretical construct or as a simulation. For the former, a set of assumptions about users is derived from the experience or specific pain points of the team who created the robot; however, these assumptions are not always tested and might lead to biased decisions during robot development. For example, among the target users for MiRo are children who are *not* interested in coding, despite the robot being developed to teach programming skills to children:

> *"A lot of kids that put aside coding in school are enjoying teaching a cute animal, which is what it [the robot] is intended to be. This is because we want to bring creative kids to coding and interacting with the robot, because a lot of kids are not into coding."* (Subject 9, Male)

This example shows how the robot's target user differed from the theoretical construct held by the team based on their previous experiences. Subject 9 elaborates on this: *"There was also a team that when designing [the robot] expected it only to be used for university students and not kids or the elderly, so they were not included at the time in the design."* Although the subset of target users was considered during certain stages of design, they were not fully integrated in all areas of the workflow. In this case, assumptions were made about children's low interest in learning to program that were not always investigated. Instead, this user demographic existed in their design but served as a theoretical construct with limited actual involvement. While this resulted in positive outcomes for this particular robot, *"Unintentionally, it [the robot] has been more successful than what we thought it would be."* (Subject 9, Male), this is not true for many robot companies that eventually cease operations.

Users can also be simulated using algorithms. In this case, instead of testing the robot with real users, the team can perform a series of virtual simulations to assess how a robot would behave in an interaction scenario among people. Another situation can take place when users are asked to provide feedback of a virtual, simulated robot. While the feedback from users can be valuable, their experience of interacting with a virtual robot can significantly differ than their experience interacting with a physical robot, which can bias the design and development process in ways that are not optimal for user adoption. For example, *"We developed a questionnaire for HRI researchers where they were asked what they wanted in terms of degrees of freedom for the robot."* (Subject 11, Male). However, the team quickly realized that users wanted more than what was feasible to achieve in a physical robot (compared to the virtual robot shown in the questionnaire). Thus, we observe the necessity of testing a physical robot with real people during robot development, where substantial changes can be made, if necessary, to avoid biases about users needs.

4.3 Team Composition When Designing Social Robots

Several main topics emerged when discussing the composition of teams that create robots: 1) all interviewed subjects belonged to *interdisciplinary teams*, 2) the majority of teams used *outsourcing* for special skill acquisition, 3) most teams relied on *international sites* to manufacture scalable robots. This section describes these topics and discusses the human dynamics underlying the challenges and success of these teams.

4.3.1 Interdisciplinary Teams

To create a social robot is to create an artificial being. Therefore, the design, development, and testing of social robots calls for *interdisciplinary team composition*. Reviewing our subjects' backgrounds (**Table 1**), we see that teams are generally composed of mechanical and electrical engineers, computer scientists, psychologists, and artists.

When referring to how their team is composed, it was mentioned: *"We had the two founders and CEO, so they have* *an electrical engineering background and a design mechanical engineering background. We had a mechanical engineer, a very serious developer (really top notch), a second electrical engineer, and then me. And I think we had a second 3D designer."* (Subject 10, Male). Another subject highlighted the richness of social robot design when knowledge from different fields is incorporated:

> *"I come from an animation and design background. I always worked to aesthetic principles, which is super important to robotics."* (Subject 3, Female)

This idea was further reinforced:

> *"The team brought in different insights. I brought the HRI part with the human scenarios, then there was the hardware of how to actually build it, and the designers were about how can we shape it. They knew what really looks good."* (Subject 10, Male)

All interviewees mentioned they had interdisciplinary teams and acknowledged the complexity of creating a social robot. The main insight here is that *interdisciplinary teams have the required knowledge to create social robots.*

4.3.2 Outsourcing Special Skills

Despite the necessity of working with interdisciplinary team members, not all team members are needed at all stages of robot development, and some roles are outsourced. This brings the advantage of decreasing the complexity of the 'core team' of robot builders and of making the product scalable and successful: *"The most robust robot that we built was with a collaborator, who was a mechanical engineering consultant."* (Subject 12, Male).

According to the interviewees, the design of the robot is explored within a small and cohesive team, and the product is then outsourced to be manufactured at scale when there is a final prototype: *"We worked with an external manufacturer; you need to come to them with a product, and they do design for manufacturing."* (Subject 1, Male). The important aspect is to provide a prototype that is 'manufacturable,' a complex topic that depends on *"all the processes that need to happen to make something at scale."* (Subject 1, Male).

It is important that external manufacturing companies have previous experience with building social robots or some type of technology: *"They should have done a robot before."* (Subject 9, Male) because of reliability issues:

> *"It's a very different 'animal' if you've built hardware but without moving parts or when you actually need to move. Here, you get into the reliability issues and how you build something that does not hurt the user but at the same time is robust enough. You need to find someone that has experience with hardware, robots, or mechatronics products somehow."* (Subject 1, Male)

Besides using external manufacturing companies to build the robot, other team roles were outsourced, such as artists, *"We had several contracted animators who also helped with designing and animating some eyes."* (Subject 4, Male); public relations team members, *"This is related to how you interact with customers; that's pretty important, and you will want to keep this in mind as well."* (Subject 1, Male); and marketing, *"Marketing dealt with public relations, mostly. We had a third-party public relations firm that worked with us a lot."* (Subject 7, Female). There were some discussions about what types of team roles should not be outsourced; for example, according to Subject 1 (Male), *"I recommend not outsourcing anything that is software; that's something you have to be building, and it is important to be in-house."*

4.3.3 International Personnel for Manufacturing

With few exceptions, the majority of the teams hired external manufacturing companies to build social robots: *"Our internal hardware team did some electrical and mechanical engineering, but we also interfaced with the various contractors. Flew to China a whole lot."* (Subject 7, Female). According to Subject 10 (Male), *"We actually had a very good relationship with a manufacturing plant in China, and they brought another piece of, like, what can you actually build and scale."* The idea that building a social robot requires joining the forces from multiple disciplines is highlighted in this quote:

> *"We're working with The Netherlands, Germany, the States, Australia, Japan, China. I mean, it's an international project."* (Subject 3, Female)

Some companies had no problem working with international teams, *"Japan, Taiwan, we worked with them to build it. The distance was not a problem."* (Subject 9, Male). Other teams struggled to find the balance between team cohesion and long-distance professional relationships:

> *"While it is great to be working across distances, it was rather difficult to coordinate the development of this robot, especially due to differing time zones."* (Subject 4, Male)

This was supported by others:

> *"I was never in China myself, but my impression of that is that to get things the way we wanted them, there had to be that very tight on the ground interaction [between the external manufacturing company and the core team]. To have someone there, keeping eyes on things and stop(ping) it from going in the wrong direction."* (Subject 10, Male)

The main challenge with international teams was to develop solid and sustainable relationships: *"It was hard to organize times for direct communication through video conferences, and getting timely responses before deadlines was difficult."* (Subject 4, Male). Inter-team synchronicity and mutual understanding was important because *"Manufacturing companies have their own team, and we want someone representing our company sitting in the meeting."* (Subject 10, Male) so that both views are represented.

TABLE 2 | Synthesis of recommendations when designing new social robots.

Design process	Role of the user	Opportunities	Obstacles
Iterative	User is part of the workflow	Flexible design process consists of improvement loops in the robot considering users' feedback	Design process can be chaotic, time consuming, and requires access to multiple users
Linear	There is minimal user involvement	Design process is well-defined and concrete. This can lead to faster results since it is easy to define cost, stages, and time	Design process is rigid, which can lead to undesirable results as it lacks iteration
Data points	User is a theoretical construct	This is an economical design process since it leverages accumulated expert knowledge or generalization	Risks include stereotyping the user and excluding non-traditional populations

Different strategies were used by the teams to build a coordinated relationship across the globe. One way teams tried to synchronize was to go on-site to the manufacturing company, *"They [the engineers] would frequently go to the factory for . . .weeks at a time when some [robot] units were coming off. They would . . . just go to China for . . .3 weeks at a time and . . .stay there and intensively watch stuff roll off the line and tweak the process."* (Subject 6, Male). However, there can be instances where a team member going on-site is impossible, so another solution was raised: *"We'd probably want to have at least an independent agent in China. That could be either someone in your own staff or you need to hire someone in China who reports to you and not the manufacturing company."* We note that most of the external manufacturing companies referred during this study were located in China chiefly because they have the *"skills and knowledge for making things super for us, like, really lowering the cost."* (Subject 12, Male).

5 DESIGN IMPLICATIONS

Throughout this paper, we exposed different approaches to creating robots. Given the knowledge gathered, we now synthesize our findings into design recommendations for new robots. These recommendations map different design processes to different ways users are brought in the design of social robot. Additionally, we elaborate on the opportunities and challenges of the approaches (see **Table 2**). We hope to inspire future robot creators to alter their design pipelines accordingly.

5.1 Robot Design With Human in Mind

Throughout this paper, we surfaced the benefits of users' involvement in the design of social robots. We consider it equally important to explicitly voice the negative impacts that *non-human-centric* robot design processes can have. A *non-human-centric design process* refers to minimal or non-existent user/customer involvement in the process of robot design. Designing robots without users in mind can lead to stereotype propagation, creation of erroneous assumptions about what the users need are, over-generalization and misinterpretation of problems, and other forms of bias, such as the creation of solutions that do not fit the user's ecosystem (Benjamin, 2019). If robots are not designed with humans in mind, they can rarely succeed in helping to solve a real need or problem, falling short in the market since consumers avoid investing in expensive products (such as robots) that do not help them in concrete ways. In this work, we argue that one of the most powerful ways to counteract biases in social robot design is to follow design justice practices by creating a design pipeline that is human-centered (Costanza-Chock, 2020). By doing so, robot creators can translate human values, voices, and needs into actionable design decisions for the robotic products they are creating.

5.2 Distinctiveness of Social Robot Design

Designing social robots is a unique process that may not apply to other technologies. Many theories and practices from the field of human-computer interaction (HCI) need to be considered in this process. For example, designing robots underlies an iterative process, based on human needs, that requires technical precision. While designing social robots shares aspects with HCI, there are unique features of this design process that are specific to HRI. A key insight from this work is that building a robot is, in some ways, equivalent to the complexity of building an artificial being. When designing social robots, we must account for variables such as robot personality, artificial emotion expression, conversational abilities, and movement/gestures. For example, a robot can have varying levels of expressivity; they are actuated and communicate through movement to change the physical world we live in; and they are almost always anthropomorphized to a certain degree. Thus, when determining a robot's "purpose," it is essential to consider the combination of these variables, which cause problems unique to social robot creation.

5.3 Variable Alignment for Successful Robots

While it was not the goal of this qualitative study to identify and correlate variables associated with successful/unsuccessful robot products, we highlight different processes and strategies used by robot creators to better understand the challenges of building robots. Besides defining the purpose of the robot, one of the most important variables for success, our work showed a set of other variables that must be aligned. These variables include working in interdisciplinary teams, relying on outsourced labor to scale the product across time, and establishing effective international professional relations. A deeper understanding of these variables made clear that robot success in the consumer market is related not only to units of sale, but to the alignment of a complex set of variables that come into play long before the robot first appears on market shelves.

5.4 Additional Aspects on Social Robot Design

This work uncovered additional information about the design process of social robots. Despite going beyond the proposed research questions, we include this information as it can influence how future robot creators conceive designing a robot.

5.4.1 Success Through Purpose

It is essential to have a concrete answer to the question, "Why are we building this robot?" As a participant mentioned, *"You should start with a need and provide a solution. You can say 'I have a robot' and then look for the problem you want to solve, but I don't think this strategy is effective."* (Subject 1, Male).

Being able to quantify how successful a robot is important and directly tied to the robot's purpose. In a university lab, success might be measured by the learning gains of students or the publication record using the created robot. In a market context, success might be evaluated per unit sales or number of customer complaints. Defining metrics of success lets us circle back to the question of "Why are we building this robot?" and evaluate whether the initial purpose for robot creation is being successfully met.

5.4.2 For Accuracy, Double Everything

In a study that evaluated time predictions for a coding task, results showed that programmers take 1.5x more time than initially expected, showing how "We are the worst at completing a task in the originally planned amount of time." (Brauer, 2021). Our study shows that this is true even for expert robot creators. As one participant mentioned, *"It sounds simple, it is like, oh yeah everything is here, we have all the things, but like getting something to actually work and be reliable for a long time and also keeping the knowledge of the complex system if they don't document, test things very, very rigorously, which is next to impossible."* (Subject 2, Male).

A major pain point identified during our study was the underestimation of effort and cost inherent in creating a social robot, which can lead a company to fail in meeting important deadlines and initially agreed upon business goals. As a participant mentioned, to be successful *"double is the metric. Double everything: time, costs, everything."* (Subject 11, Male). to address this design implication, it is crucial to identify the exact robot functionalities to combat the tendency to add unplanned features. In social robot design, it is easy to lose track of the initial vision for the robot when new insights and feedback are being delivered by users testing it. We argue that instead of designing for features, robot builders could adopt the approach of *designing for meaning*. Towards this end, users' preferences should drive robot design decisions in meaningful ways, keeping in mind the original purpose for the robot.

5.4.3 Simplicity in Design, Robustness in Function

The combination of simplicity and robustness in a robot are two design values that matter for its success. As participants mentioned, *"If I had to build a robot, I'd build a waaaay less complex and smaller robot."* (Subject 8, Female), and *"It is preferable [to build a robot] that is simple, niche, and that can be done well."* (Subject 1, Male). Simplicity has been a core design principle adopted in many ways in the design of technologies (Chang et al., 2007); it avoids anything getting in the way of the user and is defined in design as the *lack of obstruction* (Karvonen, 2000). However, most current robots fall into the category of humanoid robots, with highly complex features that run counter to the principle of simplicity. Additionally, humanoid robots frequently fall prey to the *uncanny valley* effect, i.e., feelings of uneasiness towards a robot that looks like a human but is not really human, which could be avoided by taking a simpler design approach that allows for more robustness (Mori et al., 2012).

Striving for simplicity, not only in terms of hardware but also in terms of robot identity and behavior, seems important when creating a robot. In this sense, the robot should pass a clear message to the user about what it is, what it can do, and how it will behave. Given this work, we argue that this can be conveyed in a robot through its aesthetics, such as its materiality, colors, and motions.

6 CONCLUSION

This work shed light on the design, development, and testing processes of creating social robots. The main goal of this qualitative investigation was to provide in-depth insights about building robots as marketable products that work in the real world. We have shown the existence of several layers in the design of robots: from different development processes, to several degrees of user involvement, to the complexity of team compositions. All things considered, creating robots is an extremely complex process that requires the alignment of many variables to result in a successful and lasting market product. It is thus important to question and consider the value of a robot in our lives and its place in the socio-technological world we live in and the future we want to create.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because this was not part of our IRB. Requests to access the datasets should be directed to PA-O, patri@uw.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB from the University of Washington.

AUTHOR CONTRIBUTIONS

PA-O conducted the interviews. PA-O and AO analyzed the data and wrote the paper. PA-O, EB, and MC designed the study and discussed results. EB and MC supervised this work.

FUNDING

This work was supported with funding from the Innovation Gap Fund from CoMotion at the University of Washington and by the NRI: INT EMAR Project.

ACKNOWLEDGMENTS

We are grateful to the robot creators we interviewed and talked to for revealing more about their experience in creating social robots.

REFERENCES

Axelsson, M., Oliveira, R., Racca, M., and Kyrki, V. (2021). Social Robot Co-design Canvases: A Participatory Design Framework. *ACM Trans. Human-Robot Interact. (THRI)* 11, 1–39. doi:10.1145/3472225

Baraka, K., Alves-Oliveira, P., and Ribeiro, T. (2020). "An Extended Framework for Characterizing Social Robots," in *Human-Robot Interaction* (Springer), 21–64. doi:10.1007/978-3-030-42307-0_2

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.

Björling, E. A., and Rose, E. (2019). Participatory Research Principles in Human-Centered Design: Engaging Teens in the Co-design of a Social Robot. *Multimodal Technol. Interact.* 3, 8. doi:10.3390/mti3010008

Bocken, N. M. P., De Pauw, I., Bakker, C., and Van Der Grinten, B. (2016). Product Design and Business Model Strategies for a Circular Economy. *J. Industrial Prod. Eng.* 33, 308–320. doi:10.1080/21681015.2016.1172124

Brauer, K. (2021). *i'll Finish it This Week" and Other Lies. arXiv preprint arXiv: 2103.16574*.

Buchanan, R. (2001). Human Dignity and Human Rights: Thoughts on the Principles of Human-Centered Design. *Des. issues* 17, 35–39. doi:10.1162/074793601750357178

Campbell, J. L., Quincy, C., Osserman, J., and Pedersen, O. K. (2013). Coding In-Depth Semistructured Interviews. *Sociol. Methods & Res.* 42, 294–320. doi:10.1177/0049124113500475

Chang, A., Gouldstone, J., Zigelbaum, J., and Ishii, H. (2007). "Simplicity in Interaction Design," in *Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, 135–138. doi:10.1145/1226969.1226997

Chun, B., and Knight, H. (2020). The Robot Makers. *J. Hum.-Robot Interact.* 9, 1–36. doi:10.1145/3377343

Cooley, M. (1999). "Human-centered Design," in *Information Design*, 59–81.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. The MIT Press.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., and Sheikh, A. (2011). The Case Study Approach. *BMC Med. Res. Methodol.* 11, 100–109. doi:10.1186/1471-2288-11-100

Dautenhahn, K. (2007). Socially Intelligent Robots: Dimensions of Human-Robot Interaction. *Phil. Trans. R. Soc. B* 362, 679–704. doi:10.1098/rstb.2006.2004

Dearnley, C. (2005). A Reflection on the Use of Semi-structured Interviews. *Nurse Res.* 13, 19–28. doi:10.7748/nr2005.07.13.1.19.c5997

Elprama, S. A., Jewell, C. I., Jacobs, A., El Makrini, I., and Vanderborght, B. (2017). "Attitudes of Factory Workers towards Industrial and Collaborative Robots," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 113–114. doi:10.1145/3029798.3038309

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. *Robotics Aut. Syst.* 42, 143–166. doi:10.1016/s0921-8890(02)00372-x

Garrison, D. R., Cleveland-Innes, M., Koole, M., and Kappelman, J. (2006). Revisiting Methodological Issues in Transcript Analysis: Negotiated Coding and Reliability. *internet High. Educ.* 9, 1–8. doi:10.1016/j.iheduc.2005.11.001

Gasson, S. (2003). Human-centered vs. User-Centered Approaches to Information System Design. *J. Inf. Technol. Theory Appl. (JITTA)* 5, 29–46.

Gill, P., Stewart, K., Treasure, E., and Chadwick, B. (2008). Methods of Data Collection in Qualitative Research: Interviews and Focus Groups. *Br. Dent. J.* 204, 291–295. doi:10.1038/bdj.2008.192

Guest, G., Bunce, A., and Johnson, L. (2006). How Many Interviews Are Enough? *Field methods* 18, 59–82. doi:10.1177/1525822x05279903

Hartson, R., and Pyla, P. S. (2012). *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Elsevier.

Hoyle, R. H., Harris, M. J., and Judd, C. M. (2002). *Research Methods in Social Relations*.

Jiménez, C., Rusu, C., Roncagliolo, S., Inostroza, R., and Rusu, V. (2012). "Evaluating a Methodology to Establish Usability Heuristics," in Proceeding of the 2012 31st International Conference of the Chilean Computer Science Society, Valparaiso, Chile, 12-16 Nov. 2012 (IEEE), 51–59.

Karvonen, K. (2000). The Beauty of Simplicity. *Proc. 2000 Conf. Univers. Usability*, 85–90. doi:10.1145/355460.355478

Kawakita, J. (1991). *The Original Kj Method*, 5. Tokyo: Kawakita Research Institute.

Krippendorff, K. (2009). *The Content Analysis Reader*. Thousand Oaks, CA: Sage.

LLC, B. R. (2020). "Social Robots: Emotional Connection and Task Engagements," in *BCC Research Special Report: CON010A*, 1–64.

Moni, S. M., Mahmud, R., High, K., and Carbajales-Dale, M. (2020). Life Cycle Assessment of Emerging Technologies: A Review. *J. Industrial Ecol.* 24, 52–63. doi:10.1111/jiec.12965

Mori, M., MacDorman, K., and Kageki, N. (2012). The Uncanny Valley [from the Field]. *IEEE Robot. Autom. Mag.* 19, 98–100. doi:10.1109/mra.2012.2192811

Newcomer, K. E., Hatry, H. P., and Wholey, J. S. (2015). "Conducting Semi-structured Interviews," in Handbook Of Practical Program Evaluation *492*.

Norman, D. (2013). *The Design of Everyday Things: Revised and*. expanded edition. Basic Books.

Park, J., and Park, M. (2016). Qualitative versus Quantitative Research Methods: Discovery or Justification? *J. Mark. Thought* 3, 1–8.

Patton, M. Q. (2005). *Encyclopedia Of Statistics in Behavioral Science. Qual. Res.* doi:10.1002/0470013192.bsa514

Queirós, A., Faria, D., and Almeida, F. (2017). Strengths and Limitations of Qualitative and Quantitative Research Methods. *Eur. J. Educ. Stud.* 3. doi:10.5281/zenodo.887089

Sanneman, L., Fourie, C., and Shah, J. A. (2021). The State of Industrial Robotics: Emerging Technologies, Challenges, and Key Research Directions. *Foundations and Trends® in Robotics* 8 (3), 225–306. doi:10.1561/2300000065

Sauppé, A., and Mutlu, B. (2015). "The Social Impact of a Robot Co-worker in Industrial Settings," in Proceedings of the 33rd annual ACM conference on human factors in computing systems, Seoul, April 18–23, 2015 (Seoul: ACM), 3613–3622. doi:10.1145/2702123.2702181

Shibata, T. (2004). An Overview of Human Interactive Robots for Psychological Enrichment. *Proc. IEEE* 92, 1749–1758. doi:10.1109/jproc.2004.835383

Sofaer, S. (1999). Qualitative Methods: what Are They and Why Use Them? *Health Serv. Res.* 34, 1101–1118.

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: sage.

Still, B., and Crane, K. (2017). *Fundamentals of User-Centered Design: A Practical Approach*. Boca Raton, FL: CRC Press.

Story, M. F. (2001). "Principles of Universal Design," in *Universal Design Handbook*. 2nd Edn (New York, NY: McGrawHill).

Thompson, D. V., and Norton, M. I. (2011). The Social Utility of Feature Creep. *J. Mark. Res.* 48, 555–565. doi:10.1509/jmkr.48.3.555

Tongco, M. D. C. (2007). Purposive Sampling as a Tool for Informant Selection. *Ethnobot. Res. App.* 5, 147–158. doi:10.17348/era.5.0.147-158

Welfare, K. S., Hallowell, M. R., Shah, J. A., and Riek, L. D. (2019). "Consider the Human Work Experience when Integrating Robotics in the Workplace," in Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 11-14 March 2019 (IEEE), 75–84. doi:10.1109/hri.2019.8673139

Wurhofer, D., Meneweger, T., Fuchsberger, V., and Tscheligi, M. (2018). "Reflections on Operators' and Maintenance Engineers' Experiences of Smart Factories," in Proceedings of the 2018 ACM Conference on Supporting Groupwork, January 7–10, 2018 (Sanibel Island, FL: ACM), 284–296.

Yanco, H. A., and Drury, J. (2004). "Classifying Human-Robot Interaction: an Updated Taxonomy," in Proceeding of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), The Hague, Netherlands, 10-13 Oct. 2004 (IEEE), 3, 2841–2846. doi:10.1109/ICSMC.2004.1400763

Yildirim, N., McCann, J., and Zimmerman, J. (2020). "Digital Fabrication Tools at Work: Probing Professionals' Current Needs and Desired Futures," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, April 25–30, 2020 (Honolulu, HI: ACM), 1–13. doi:10.1145/3313831.3376621

Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*. Thousand Oaks, CA: Sage publications.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership