Hardships in the Land of Oz: Robot Control Challenges Faced by HRI Researchers and Real-World Teleoperators

Alexandra Bejarano, Saad Elbeleidy, Terran Mott,
Sebastian Negrete-Alamillo, Luis Angel Armenta, and Tom Williams

***Abstract*— Wizard-of-Oz (WoZ) is one of the most widely used experimental methodologies across the field of Human-Robot Interaction (HRI), making WoZ teleoperation interfaces a critical tool for HRI research. Yet current WoZ teleoperation interfaces are overwhelmingly tailored towards a narrow set of HRI interaction paradigms. In this work, we conducted a set of interviews with HRI researchers to better understand the diversity of teleoperation needs across the HRI community. Our analysis highlighted (1) human challenges, with respect to wizards' expertise, the need for quick responses, and research participants' unpredictability; (2) robot challenges, with respect to robot malfunctions, delays, and robot-driven complexity, and (3) interaction challenges, with respect to researchers' varying control requirements and the need for precise experimental control. Moreover, our results revealed unexpected parallels between the experiences of HRI researchers and real-world teleoperators, which open up fundamentally new possibilities for future work in robot control interfaces and encourage radically different perspectives on what types of interfaces are even needed to best facilitate WoZ experimentation. Leveraging these insights, we recommend that WoZ interfaces (1) be designed with extensibility and customization in mind, (2) ease interaction management by accounting for unpredictability and multi-robot interactions, and (3) consider WoZ teleoperators beyond the context of experimentation.**

INTRODUCTION

Wizard-of-Oz (WoZ) is commonly used by Human-Robot Interaction (HRI) researchers to remotely control robot speech and movement during research studies [1], [2]. WoZ is often used in two key ways [3]: (1) *perceptual WoZ* in which a "wizard" (the person controlling a robot) provides oracle knowledge for a particular robot perceptual capability (e.g., speech recognition) and (2) *cognitive WoZ* in which a wizard hand-controls robot decision making or action execution. WoZ enables experimenters unique experimental control, evaluation of prototyped designs, and evaluation of not-yet-feasible robot designs. Accordingly, WoZ has proved to be an effective way to quickly, easily, and safely investigate robot behaviors and human interactions with robots.

Despite the utility of WoZ, researchers face a number of challenges in actually using it. As Rietz et al. [4] indicate, there is a current lack of general robot control tools that are publicly available and easily adaptable to various robot study domains and researcher needs. Furthermore, for robots like Misty[1] and Stretch[2], out-of-the-box control interfaces may limit the degree of control a researcher can achieve. Accordingly, substantial effort is often needed to build control interfaces that actually meet experimenter needs. This raises a key question: how can Wizard-of-Oz interfaces be improved to better meet the needs of researchers?

To answer this question, we conducted a set of interviews with HRI researcher so as to better characterize the nuanced challenges faced by HRI researchers using the WoZ paradigm. Our results revealed key parallels between the robot control experiences of HRI researchers and real-world teleoperators, especially in terms of the challenges both groups face. As we will show, identifying these parallels suggests key unexplored directions for future work in robot control interfaces, and encourages a radically different perspective on the types of interfaces needed to facilitate WoZ experimentation. In response to these identified challenges and research parallels, we present design recommendations for future robot control interfaces that can better support researchers' needs for fast, reliable, customizable systems, and examples of how to enact these recommendations.

RELATED WORK

As analyzed by Riek et al. [2], WoZ interfaces are typically used to control one or more of six key dimensions of robot perception and cognition: (1) *natural language processing* (e.g. speech understanding and output [5], [6]), (2) *nonverbal behavior* (e.g. head movement, gaze [7], [8]), (3) *navigation and mobility* (e.g. positioning [9], [10]), (4) *manipulation* (e.g. grasping [11], [12]), (5) *sensing* (e.g. seeing [13], [14] capabilities), and (6) *mapping and localization* (e.g. knowing where a robot is placed [15]).

WoZ interfaces enable wizards to control any number and combination of these robot capabilities to overcome key robot limitations [2], [16]. As such, the WoZ method can help researchers investigate robot behaviors and human-robot interactions without needing to implement fully autonomous systems. Additionally, this method can help wizards maintain certain safety precautions to protect participants or bystanders from unexpected or inappropriate robot behaviors. For example, studies with vulnerable populations may require human expert supervision and intervention to ensure best practices and safety that may not be possible to ensure when using autonomous robots (e.g. the use of socially assistive robots with children [17], [18] or older adults [19]). Due to its benefits, WoZ has been used in a variety of work such as in conducting therapy and education sessions with children [17], [20], [21], advancing the development and

The authors are with the MIRRORLab at the Colorado School of Mines, Golden, CO, USA, {abejarano,selbeleidy,tmott, snegretealamillo,armenta,twilliams}@mines.edu

[1]mistyrobotics.com

[2]hello-robot.com

interaction of autonomous vehicles [22], [23], [24], exploring how to create more natural communication between robots and humans [25], [26], and understanding how robots may influence human behavior and perception [27], [5], [28].

Yet WoZ also comes with several challenges, including concerns about its use and standardization [2], ethics due to the deception involved [29], [30] and its effects on researchers [31]. Moreover, it can be difficult for wizards to enact certain robot actions in real-time, or to consistently control preplanned robot behaviors, due to the mismatch between dynamic environments and static robot action execution [32]. These challenges are especially salient for in-the-wild [33] and longitudinal studies [34]. To make matters worse, even the process of initially setting up a robot to be controlled can require considerable work from researchers to decide which robot actions need to be manually controlled by a wizard or automated, and how specifically to implement control for wizarded actions [32]. As such, to understand the breadth of challenges today's researchers face with robot control, we conducted interviews with HRI researchers about their experiences with wizarded HRI studies.

METHODOLOGY

Interview Design and Participants

We conducted IRB-approved 1-hour interviews with participants over video chat. After providing informed consent, participants were asked questions (listed in Appendix) designed to assess their experiences with WoZ. In particular, we asked participants about their current and future research work and the robot control interfaces they might require.

We recruited six HRI researchers from other universities (three professors and three graduate students). All participants had experience conducting human-subject WoZ studies and/or investigating the design of interfaces and tools for robot programming and control within the field of HRI, demonstrating high sample specificity[3]. Tables I and II summarize participants' expertise and experience.

Analysis

Interviews were recorded via Zoom, transcribed, annotated using the Dovetail[4] qualitative analysis software, and then analyzed using a Grounded Theory methodology [36]. Three researchers performed open coding of all interview responses, yielding 467 open codes that were then clustered into axial codes and grouped into two overarching themes: HRI experimentation and robot control.

RESULTS

Participants demonstrated a range of insights about the logistics of robot control. In particular, participants discussed the capabilities of the different types of robots they used during experiments[5], and the different tasks and domains their robots were used for (Tab. I). Participants also discussed specific robot control practices, the number of operators they typically needed, the devices and programs they used, and how they prepared for teleoperation sessions (Tab. II). Overall, our analysis of these interviews revealed three types of challenges researchers faced: Challenges with (1) humans, (2) robots, and (3) interactions. In this section, we discuss each of these sets of challenges.

Challenges with Humans

Participants described several challenges that occur as a result of human involvement in research studies,relating to both robot teleoperators and experimental participants. Our participants highlighted the need for easy and quick means of robot control to address the challenges of dealing with teleoperators' non-expertise, teleoperators' need for processing time, and interactants' unpredictability.

Teleoperators are Non-Experts: P1 and P5 both described how the complexity of robot control necessitates a certain level of expertise/training. For example, P5 said: "*With the joystick controller for the fetch robot, if you wanna move the arm, it's really difficult because it doesn't make sense how to move an arm on a joystick...If I wanted to do something, you have to become like an expert with the joystick to make it do exactly what you want it to do.*" However, teleoperators may not be experts with the robots used in particular studies. As such, to minimize training time for teleoperators, who are often undergraduates, WoZ interfaces ought to be easy to learn and use by non-experts.. For instance, P1 argued that interfaces must be sufficiently user friendly so that undergraduate students can figure out how to use them. P1 described liking drag-and-drop programming methods for pre-programming robots as they do not require prior programming experience and are highly intuitive. P2 provided a first-person perspective on this: "*A lot of these interfaces I use like in a very short term user study setting. So I obviously don't have as much time to learn it as I might if I actually work with it on a regular basis.*"

Teleoperators Require Time to Process: Participants discussed the importance of teleoperators being able to quickly receive, interpret, and act on data from a robot to keep up with the pace of interactions. Yet P1, P2, P3, and P6 all suggested that current interfaces do not enable teleoperators to react to interactants with sufficient speed. For instance, P3 mentioned: "*Areas that I think are most challenging are as an operator, especially in speech oriented systems, is the fact that you often can think fast enough but you can't type fast enough and so the interaction slows down.*" Yet P1 and P3 both imagined ways to speed up teleoperator input to robots. P3 imagined automated transcription (and possibly re-synthesis of teleoperator's speech) to prevent the need to type robot speech. Similarly, P1 said: "*I would want to be able to have a number of behaviors that I might want the robot to do and to maybe categorize them or or cluster them so that when it's time I can like easily with my mental model like find the section it'll be in and click on it pretty quickly.*"

[3]Information power [35] was used to determine the quality of our interviewed sample.

[4]dovetail.com

[5]Specific robots are not listed per participant as this may deanonymize participants. Some robots mentioned include Stretch, Baxter, Create, Pepper.

ID	Role	No. Robots	Types of Robots Used	Types of Control Needed	Robot Tasks & Domains
P1	Professor	1-20	humanoid, functional. physically embodied & simulated	movement, speech, movement & speech coordination	search & rescue tasks, interactions in public spaces
P2	Graduate Student	1	functional, manipulators. physically embodied	movement	pick & place tasks
P3	Professor	1	embodied in existing things, minimal. physically embodied & simulated	speech, movement, movement coordination, lights, music, graphics	interactions in homes & in vehicles, delivery
P4	Professor	1-6	social, custom-built, telepresence, non-anthropomorphic. physically embodied	movement, speech, non-verbal sound cues (beeps, music, motor sounds)	service tasks, interactions in homes
P5	Graduate Student	1	mobile, manipulators. physically embodied & simulated	mobile movement, arm manipulation, gaze	pick up & move items, search buildings, household tasks
P6	Graduate Student	1	mobile manipulator. physically embodied	movement, manipulation, speech	interactions in healthcare, pick up & deliver items

TABLE I: Summary of Participant Robot Usage

ID	No. of Operators per Robot Needed	Devices & Programs Used	Operation Prep & Guidance
P1	aims for 1 operator per robot due to scheduling conflicts	mouse & keyboard, commandline, Choregraphe	uses detailed document with protocol & troubleshoot instructions to assist operators. operators need to practice protocol beforehand
P2	1 operator per robot	mouse & keyboard, AR headset, custom browser interface, commercial software, Unity	uses written reference material & experimenter demos (in-person & videos) to train operators. creates scripts to initialize study operation
P3	2 operators per robot (1 navigation, 1 interactivity)	mouse & keyboard, joystick, on-screen controls, custom browser interface, Unity	uses simulated studies for operators to train & practice beforehand. prescripts common comments & phrases
P4	typically 1 operator per robot. 1 operator has controlled up to 6 robots	mouse & keyboard, gamepad controller, custom browser interface, commandline, Adobe Audition, Keynote	preplans & uses protocol document to guide operators on setup & study operation
P5	prefers 2 operators per robot (1 control, 1 safety supervision), but usually only 1 operator per robot	mouse & keyboard, joystick, VR/AR headset, custom built interface, commandline	preplans & creates prewritten scripts to run during study operation
P6	2 operators per robot (1 movement, 1 speech)	mouse & keyboard, gamepad controller, custom browser interface, commandline, commercial software	creates launch files to help initialize robot & controls

TABLE II: Summary of Participant Control Logisitics

Interactants are Unpredictable: P6 described the difficulty of pre-planning robot behaviors, especially for in-the-wild and open-ended studies in which it is difficult to predict interactants' interactions with a robot. For example, P6 described their difficulty predicting how *humans* would act in open-ended contexts: "*I think because the work that we do is so open-ended, like it's not easy to script things because we just can't really predict what people will say or do because we're doing like in the wild kind of open-ended design problems.*" As P6 further noted, even in situations where there is an established researcher-participant relationship for in-the-wild studies (e.g., in which the researcher is familiar with prior participant interactions with a particular robot), it can still be difficult to "*predict what the robot needed to say or do in advance. So it's also somewhat of like just on the fly having to add things as necessary.*" As such, similar to the prior section, P6 wondered of ways to alleviate this challenge and potentially ease the workload of teleoperators: "*If some like AI assistance could like help us like you know, maybe just based on what that person is saying, if we could like offer some suggestions for some potential responses then we could just kind of quickly edit.*"

Challenges with Robots

Participants highlighted that consistency was a key requirement for robot control when conducting research studies especially to ensure robots do not make errors and can continuously do what is intended. However, consistency can be hard to maintain during studies due to robot-specific challenges. In particular, participants pointed out challenges with robot malfunctions and delays. Participants also discussed how these existing challenges are then amplified in multi-robot interactions due to the increased number of robots and the potential for complex identity configurations.

Robots Malfunction: All participants described unexpected technical difficulties during robot control. P1 and P5 highlighted the challenge of errors stemming from inconsistent robot actions. For instance, P1 noted: "*They don't do the thing, they do the thing and don't stop doing the thing, they do something other than the thing.*" Additionally, all participants but P2 highlighted hardware and networking challenges encountered during studies, like losing connection to a robot's camera (thus losing visual feedback from a robot) or the internet (thus being unable to communicate commands to a robot), and even having difficulty establishing a network connection to a robot in the first place. P6 recalled that: "*Every once in a while the camera would disconnect and that was very inconvenient. So we kind of had to, you know, snoop into the room a little bit to see what's going on and hope that we're maintaining our cover.*"

Participants also discussed how such malfunctions make it difficult for them to pre-plan robot behaviors for studies due to the uncertainty on whether robot actions are to succeed as expected. For instance, P1 indicated: "*[The robots] had a couple of behaviors we could do...but we never programmed them in advance to do any of those because the situation was too complex to program them in advance and expect it to work and, I did not know how to do that with these robots.*" Similarly, P5 described: "*It's annoying getting a script exactly how I want it because it's hard to kind of debug on a robot since robots, you know, react in the real world in real time and you don't know if something's gonna break until it actually breaks in real time.*"

Robots have Delays: In addition to the importance of teleoperators being able to quickly receive, interpret, and act on data from the robot (as discussed above, participants also highlighted the importance of the robot being able to quickly receive, interpret, and act on their input. Yet P1, P3, and P6 highlighted issues with delays that regularly hinder human-robot interactions, including delays in incoming data/video feed from a robot and delays between when a command is given and when a robot acts on that command. For instance, P6 recalled a moment when delay entirely prevented certain robot actions: "*It's just not possible to have a real time handoff with the robot and [a human] when you have a ten second delay in the camera feed. So I think just being able to get the information like very quickly and then also being able to have the robot respond very quickly in both the speech and the movement [are] very important.*" P1 also noted: "*[Processing delays] may make it harder to find significant results if it's inconsistent what the robots are doing...and sometimes it's like man I wish it took less time than the response time of the human because it's pretty awkward having a conversation. The robot's pausing all the time.*"

More Robots More Complexity: While only two participants explicitly mentioned working with multiple robots, all participants imagined future work involving multiple robots, such as a robot arm and social robot simultaneously interacting with a human, collaboration between home robots and IOT devices, fleets of delivery/sidewalk robots, robots collaborating to perform household tasks like moving furniture, and teams of caregiver robots in healthcare settings. These discussions showed the extent to which HRI researchers are thinking about multi-robot contexts for future HRI and considering the potential challenges of those contexts.

P1, P2, and P5 all mentioned the increased difficulty of controlling multiple robots. P1, who has worked with multiple robots, said: "*A lot of the times that I've worked with multiple robots in human-robot studies, I've made the multiple-robot-with-human interaction very brief because it's so much...And of course with multiple robots you just multiply the possibilities that you'll have something go wrong.*" P4, who has also worked with multiple robots, noted the multiple *other devices and programs* that would be needed for multi-robot control and the need to centralize multi-robot control: "*I keep talking about these disparate parts needed to control robots...Sometimes there are multiple physical laptops and multiple windows open on each one. So having a way of centralizing that would be amazing.*" Meanwhile, P5, who hadn't yet worked with multiple robots, said: "*It's hard enough getting one robot to do what you want and then to have another one that's coordinating with that robot and a human is just exponentially more difficult. And so right now I don't see an easy way of doing that.*"

Furthermore, participants also highlighted the ways that robots with different identities or personalities increase the complexity of multi-robot control. P1 and P4 described how a robot identity may need to change or migrate across robot bodies. For example, P1 said: "*You've got plenty of different things you can manipulate in terms of multiple personalities for one body, multiple personalities for multiple bodies, the same personality for multiple bodies. [There may be] like ten bodies and five personalities so that each of them has to share some, maybe some change personalities that are shared with certain others but not others. Like there's so much complexity that could go into it, especially as we're moving more towards actually caring about studying groups of robots.*"

Challenges with Interactions

As shown in Table I, participants indicated a variety of tasks and domains in which they used robots. Due to this variety, participants indicated several challenges that stem from having specific interaction needs within their particular tasks/domains. Specifically, participants pointed to challenges with interactions requiring specific types of robot control and precise control while not having the proper interfaces to readily facilitate those interactions.

Researchers Require Varied Modalities of Control: The wide array of multi-modal channels needed during robot control in interactions presents a distinct challenge, not only due to the need to synchronize timing across channels, but also due to the lack of extensibility of current robot control interfaces to account for multiple and/or new modalities. Several participants expressed frustration with the lack of control provided by existing software and indicated that to meet their particular interaction needs, they often had to create their own custom interfaces. For example, P6 highlighted the inability to control anything but motion using existing software: "*It did not incorporate the speech element. It was more about just like moving the robot and because we also needed that speech element...I could not figure out how to modify it so we just had to make our own.*" As such, participants speculated about different ways WoZ interfaces might be modified to facilitate extensibility to new modalities. For instance, P6 expressed a desire to change the speed at which a robot moves, or synchronize robot speech and movement within a single interface.

Interactions Require Precise Control: Participants also indicated that to appropriately facilitate certain interactions, precise robot control was needed. However, precise robot movement can be difficult to control and may hinder teleoperation. For example, P2 said: "*I think the biggest challenge is precise manipulation. Like getting the gripper aligned appropriately. That's even difficult [when] I'm actually there in person trying to get the robot like aligned by hand. But if I do it through an interface it can be quite difficult.*" As such, some participants speculated about ways WoZ interfaces could automate the details of command execution to enhance the speed and precision of teleoperation within particular interactions. Specifically, P1 and P6 both speculated about how future systems could accept higher-level, non-explicit commands for both speech and movement. For instance, P1 said: "*If you could say here's what I generally want them to say and then have the robots like be in really close agreement versus very distant agreement.*" Similarly, P6 imagined specifying robot waypoints to "*Take a little bit of the load off of the person driving the robot around.*"

DISCUSSION

Across our interviews, our participants highlighted the series of challenges they face while controlling robots. Our results suggest that robot control interfaces need to be designed to better support researchers' need for speed, reliability, and customizability. In interpreting these results, we were surprised to note the extent to which our results converged with other extremely recent research on robot teleoperation. Specifically, we noted several parallels to work on teleoperated Socially Assistive Robots (SARs) already being used in the real world by stakeholders like therapists and educators [18]. Comparing these two user groups revealed consistent insights that strengthened our understanding of WoZ interface design needs. As such, we will begin by describing the parallels we identified between WoZ in HRI experiments and SAR teleoperation in-the-wild.

Real-World Teleoperators Face Similar Challenges as HRI Researchers

While HRI researchers often use Wizard-of-Oz teleoperation as a prototyping or testing method, WoZ is also being used to teleoperate Socially Assistive Robots (SARs) in the wild by stakeholders across domains like healthcare, education, and therapy [37], [38]. In these domains, teleoperation is advantageous because it keeps power in the hands of human experts who can competently adapt to ethically and emotionally sensitive interactions [39]. Researchers have studied how SAR teleoperators adapt to the challenges of using robots [18] and how to improve the design of control interfaces for these user communities [20].

The challenges faced by these teleoperators in domains like therapy and education have striking similarities to the challenges faced by HRI researchers that we identified in this work. In both domains, for example, teleoperators must respond promptly to maintain the pace of conversation and adapt to unexpected environmental changes. These parallels suggest that the design guidelines proposed to improve SAR teleoperation could similarly be used to improve HRI experimental WoZ interfaces. For example, researchers have recommended certain ways that SAR teleoperation interfaces should be enhanced to give operators more control in the moment. But moreover, these parallels suggest that higher level theories and perspectives made in the context of the SAR teleoperation space might also be used to broaden our consideration of what WoZ fundamentally entails, and the broader space of WoZ teleoperators' needs that may have gone unmentioned by our participants. Specifically, researchers have found that SAR teleoperators have to perform numerous tasks "beyond the session" to effectively operate robots during sessions.

While our participants and SAR teleoperators represent user communities with significantly different objectives and levels of technology expertise, these parallels suggest that a shared set of design strategies could benefit robot teleoperators both within and beyond academia.

Teleoperators Require Speed and Adaptability: Similar to what we observed in this work, users of teleoperated SARs in the wild require fast teleoperation to keep pace with conversations, while adapting to the unpredictable nature of social interactions [18]. In both contexts, these challenges stem from technological malfunctions and delays, cognitive demands placed on operators, and the unpredictable nature of social interaction. Researchers have recommended that SAR teleoperation tools be improved to support teleoperators in adapting on-the-fly when unexpected events occur [18]. WoZ interfaces may benefit from similar improvements due to the challenge of wizarding fast, reliable interactions while compensating for unexpected technical issues and interactions.

Teleoperators Require Awareness Support: Both SAR and HRI teleoperators require a high level of awareness of what is happening during interactions to properly process, adapt, and control robots' behaviors. SAR teleoperation requires a high level of attention on all parts of an interaction including environmental factors, interactant feelings, behaviors, and goals, and robot behaviors and condition [18]. To support this required attention, researchers have recommended that SAR teleoperation tools help maintain teleoperators' situational awareness, be easy to learn and process, and aim to avoid any unnecessary complexities that may further distract teleoperators. While HRI teleoperators conducting studies may be in more controlled settings, they similarly have to keep track of such matters to maintain some level of consistency despite robot-specific challenges and interactant unpredictability. As such, WoZ interfaces may also benefit from similar aims especially when it comes to multi-robot control. For instance, as mentioned by P4 later on, control of multiple robots can be challenging as it introduces needing to keep track of even more disparate parts which could be addressed by a centralized tool.

Teleoperators Require Customization: Both SAR teleoperators and HRI researchers face a need for adaptability through customization. SAR teleoperators need to modify and organize content within a teleoperation interface (e.g. organizing content chronologically, or by utterance type). While not explicitly discussed by our participants, we imagine that HRI researchers could benefit from similar interface features due to the customization needs that our participants *did* discuss, such as the unique requirements for physical and social control that characterize interactions. As such, if WoZ interfaces afforded more flexibility in how interaction content was organized, this could similarly help to maintain quick interactions. This could address comments such as P1's expressed desire for some means of organizing robot behaviors to make interface selections easier and faster.

Teleoperation Requires Thinking "Beyond the Session": A key result of previous research on robot teleoperators is that a significant portion of the burden of using robots has to do with tasks "beyond the session", i.e., outside the context of teleoperated interactions [18]. For those operating robots in the wild, preparation, content authoring, documentation, and evaluation are all time-consuming responsibilities. These tasks happen on a much slower timeline and require more collaboration from other stakeholders. Critically, these auxiliary tasks require a fundamentally different set of interface features. While interfaces must support operators' fast, adaptable thinking during a session; interfaces are also needed that support methodical, collaborative preparation and documentation outside the session. HRI researchers have analogous needs. They must prepare content for WoZ experiments, and must document and evaluate their results. Researchers must also prepare reference material to guide novice operators and troubleshooting instructions to address unexpected errors and maintain experiment consistency. For open-ended and multi-robot interactions, these preparation and content creation tasks can be even more difficult and time consuming due to increased task complexity and variation. These insights suggest a need to fundamentally rethink the scope of WoZ teleoperators' needs: *Robot control interface designers must accommodate key requirements and responsibilities that fall "beyond the session" of individual experimental sessions.*

Taking this broader perspective helps identify the ways that specific tool development activities proposed in the SAR teleoperation space could also help HRI researchers. For example, as discussed above, the researchers designing the experiment and preparing troubleshooting content are often not the individuals who teleoperate the robots in practice. This is very similar to challenges faced in SAR teleoperation. To address these needs in the SAR teleoperation space, researchers have recommended that robot control systems include separate user interfaces for content authoring and documentation that would reduce the burden of using robots by providing opportunities for collaboration, such as content sharing and collaborative content authoring [18]. We argue that all of these enhancements could also be beneficial for researchers working together in an academic context.

To summarize, the fact such strong similarities exist across significantly different use-cases for robot teleoperation interfaces suggests that many teleoperation interface features may generalize across domains, and may improve the experiences of both user communities.

Recommendations for Future Interfaces

As our interviews showed, HRI researchers face a number of challenges during WoZ experiments that demand technical advances to robot control interfaces. Through discussion of the challenges HRI researchers face during robot control as well as the parallels between the experiences of real-world teleoperators and researchers, we derived a set of general design recommendations for HRI researchers seeking to design future robot control interfaces that meet these challenges.

Design for Extensibility and Customization: The variety of tasks and domains in which human-robot interactions may take place presents challenges to robot control that require distinct solutions on a case-by-case basis based on the particular interaction needs. As such, we recommend that **future robot control interfaces be designed for extensibility and customization**. Critically, this need for extensibility and customizability applies to both the interfaces used to teleoperate robots during experiments *and* the interfaces used to perform actions that extend beyond the session, such as content authoring. Moreover, this extensibility and customizability should be tailored both to expert and novice users. For expert users, extensibility may require software to be open-source and easily modifiable. For novice users, customizability may require interface elements to be easily reconfigured and robot behaviors to be easily re-parameterized. Overall, we envision that extensible and customizable interfaces should enable users to (1) adjust preset variables (e.g. robot speed, common speech utterances, name), (2) create duplicate interface sections for controlling additional robots, (3) modify sections with their own programming (e.g. adding movement control alongside speech), (4) adjust the format/layout of the interface (e.g. move sections to preferred areas or change colors), and (5) store multiple different session settings for easier access at later times.

Moreover, just as SAR teleoperators need to share content [18], WoZ interface configurations should be easily exportable so that they can be shared with others, including other researchers working on similar projects within a research team, as well as the broader HRI research community. Enabling extensibility *and shareability* in this way would also enhance efforts to replicate and reproduce experimental results across the HRI community.

Design for Easy Management of Interactions: The inconsistency of robot behaviors and unpredictability of human interactants presents challenges to robot control that require teleoperators to maintain a high level of situational awareness and to quickly adapt to any unexpected occurrences. As such, we recommend that **future robot control interfaces be designed for easy management of interactions**. Addressing this recommendation will require HRI researchers to consider how different interface designs specifically facilitate the tracking, managing, and controlling of robots' behaviors, data feeds, and conditions (e.g. battery life or connection stability). This recommendation is especially critical for multi-robot contexts as our interviews demonstrated how multi-robot contexts may exacerbate the challenges researchers already face with single robots due to increased complexity. As such, this would also require consideration for how an interface facilitates the tracking, managing, and controlling not only of multiple robot bodies, but also the dynamically changing associations between those bodies and different robot identities. Maintaining sensitivity to these factors will require careful attention even to how different robots are depicted within interfaces (e.g., whether an image needs to be used rather than a generic icon to better allow for quick identification of robot bodies).

Design for "Beyond the Session": The contextualization of our results through the lens of Elbeleidy et al. [18]'s recent work on SAR teleoperators helped us to see that it is necessary to consider the needs of teleoperators not just during HRI experiments, but also during other tasks necessary for conducting WoZ experiments, such as when authoring and organizing content for new experiments, or when analyzing experimental results. Put simply, we recommend that **designers of future robot control interfaces need to consider the needs of WoZ teleoperators beyond the context of the experiment.** Following this recommendation will fundamentally require researchers to rethink not only the way they develop WoZ interfaces, but moreover to rethink what types of interfaces they are even seeking to develop. We foresee a number of new technical advances that HRI researchers could make in designing new interfaces for these unexplored phases, such as content authoring. For example, generative AI tools could be integrated into interfaces designed to assist WoZ teleoperators. While the most obvious instinct might be to integrate these tools into teleoperation interfaces to provide a means of automating robot speech generation in contexts not anticipated by experimenters (a natural response given that WoZ teleoperators expressed a need for increased automation of low-level robot behaviors), this approach could actually be slower than crafting a response on-the-fly, and could be counterproductive given the need for tight experimental control. Thinking "beyond the session", however, helps us to see other ways that generative AI tools might be used without incurring those risks. For example, generative AI tools could be integrated into *WoZ content authoring interfaces* to ease the preparatory burden faced by both SAR teleoperators and HRI researchers, without meaningfully reducing experimenter control.

Alternatively, just as analysis of SAR teleoperation sessions has helped previous researchers to identify distinct types of robot speech actions needed in therapy sessions that have distinct organizational requirements and temporal dynamics [40], so too might these same categories prove useful for HRI researchers when organizing and visualizing robot dialogue options for WoZ experiments. Similarly, thinking "beyond the session" helps us to novelly consider the technical advances that might be made to facilitate HRI researchers' needs after the session. Just as SAR teleoperators need post-session analytics dashboards that summarize and visualize the actions they took during lessons and other activities [41], so too might HRI researchers need post-experiment dashboards that visualize and summarize information about an experimental session that can be readily derived from WoZ teleoperation data.

CONCLUSION

In this work, we conducted interviews with HRI researchers (n=6) to understand their experiences conducting WoZ research studies. By analyzing these interviews, we identified several challenges that researchers face during WoZ studies, including challenges relating to the humans (both teleoperators and interactants) and robots involved in studies as well as the specific needs of the interactions taking place. But moreover, we identified unanticipated connections between these challenges and the types of challenges reportedly faced by therapists and educators in Socially Assistive Robot teleoperation. Both HRI researchers and SAR teleoperators require speed, adaptability, support, and customization during robot control, yet face nuanced challenges that impede those needs. By identifying these parallels, we were able to identify a broader need to think "beyond the session", and to consider the preparatory and evaluation needs that HRI researchers likely face but may be less aware of and thus less likely to articulate.

Altogether, these research efforts allowed us to articulate key design recommendations for future robot control interfaces that can better support researchers' needs, as well as examples of how we could envision those recommendations being implemented. Specifically, we recommend that future robot control interfaces (1) be designed for extensibility and customization, (2) be designed for easy management of interactions, and (3) be designed for "beyond the session".

In future work, we aim to leverage these recommendations to design and implement a general robot control research tool that may be adapted to different research tasks and domains, as well as pre-experiment authoring and post-experiment analytics interfaces as suggested by the parallels to teleoperated SARs.

ACKNOWLEDGEMENT

This work was supported in part by ARL DCIST grant W911NF-17-2-0181.

REFERENCES

[1] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Transactions on Information Systems (TOIS)*, vol. 2, no. 1, 1984.

[2] L. D. Riek, "Wizard of oz studies in hri: a systematic review and new reporting guidelines," *Jour. Human-Robot Interaction*, 2012.

[3] P. Baxter, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme, "From characterising three years of hri to methodology and reporting recommendations," in *Int'l Conf. on Human-Robot Interaction*, 2016.

[4] F. Rietz, A. Sutherland, S. Bensch, S. Wermter, and T. Hellström, "Woz4u: an open-source wizard-of-oz interface for easy, efficient and robust hri experiments," *Frontiers in Robotics and AI*, vol. 8, 2021.

[5] E. Law, V. Cai, Q. F. Liu, S. Sasy, J. Goh, A. Blidaru, and D. Kulić, "A wizard-of-oz study of curiosity in human-robot interaction," in *Int'l Conf. on Robot and Hum. Interactive Comm. (RO-MAN)*, 2017.

[6] S. Seok, E. Hwang, J. Choi, and Y. Lim, "Cultural differences in indirect speech act use and politeness in human-robot interaction," in *Proc. Int'l Conf. Human-Robot Interaction (HRI)*, 2022.
[7] D. Rifinski, H. Erel, A. Feiner, G. Hoffman, and O. Zuckerman, "Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction," *Human–Computer Interaction*, vol. 36, no. 4, 2021.
[8] M. Shidujaman, H. Mi, and L. Jamal, ""I trust you more": A behavioral greeting gesture study on social robots for recommendation tasks," in *Int'l Conf. on Image Processing and Robotics*. IEEE, 2020.
[9] M. R. Fraune, Y. Nishiwaki, S. Sabanović, E. R. Smith, and M. Okada, "Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots," in *Proc. HRI*, 2017.
[10] K. Fischer, S. Yang, B. Mok, R. Maheshwari, D. Sirkin, and W. Ju, "Initiating interactions and negotiating approach: a robotic trash can in the field," in *2015 AAAI Spring Symposium Series*, 2015.
[11] M. Ralph and M. A. Moussa, "Toward a natural language interface for transferring grasping skills to robots," *IEEE Trans. Robotics*, 2008.
[12] N. Tran, J. Rands, and T. Williams, "A hands-free virtual-reality teleoperation interface for wizard-of-oz control," in *Proc. Int'l WS on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.
[13] M. Shiomi, T. Kanda, S. Koizumi, H. Ishiguro, and N. Hagita, "Group attention control for communication robots with wizard of oz approach," in *Proc. Int'l Conf. on Human-Robot Interaction*, 2007.
[14] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "A model of proximity control for information-presenting robots," *IEEE Transactions on Robotics*, vol. 26, no. 1, 2009.
[15] K. Fischer, "Interpersonal variation in understanding robots as social actors," in *Proc. Int'l Conf. on Human-robot interaction (HRI)*, 2011.
[16] A. Dahiya, A. M. Aroyo, K. Dautenhahn, and S. L. Smith, "A survey of multi-agent human–robot interaction systems," *Robotics and Autonomous Systems*, 2023.
[17] A. Tanevska, N. Ackovska, and V. Kirandziska, "Advantages of using the wizard-of-oz approach in assistive robotics for autistic children," in *Proc. Int'l Conf. for Informatics and Information Tech. (CiiT)*, 2016.
[18] S. Elbeleidy, T. Mott, D. Liu, E. Do, E. Reddy, and T. Williams, "Beyond the session: Centering teleoperators in socially assistive robot-child interactions reveals the bigger picture," *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, 2023.
[19] S. Thunberg, F. Angström, T. Carsting, P. Faber, J. Gummesson, A. Henne, D. Mastell, J. Mjörnman, J. Tell, and T. Ziemke, "A wizard of oz approach to robotic therapy for older adults with depressive symptoms," in *Comp. Int'l Conf. on Human-Robot Interaction*, 2021.
[20] S. Elbeleidy, T. Mott, D. Liu, and T. Williams, "Practical considerations for deploying robot teleoperation in therapy and telehealth," in *Int'l Conf. Robot and Human Interactive Comm. (RO-MAN)*, 2022.
[21] S. Ondáš, M. Pleva, and J. Juhár, "Children-robot spoken interaction in selected educational scenarios," in *Proc. Int'l Conf. on Emerging eLearning Technologies and Applications (ICETA)*, 2022.
[22] N. Martelaro, "Wizard-of-oz interfaces as a step towards autonomous hri," in *2016 AAAI spring symposium series*, 2016.
[23] K. Bengler, K. Omozik, and A. I. Müller, "The renaissance of wizard of oz (wooz)–using the wooz methodology to prototype automated vehicles," *Proc. Human Factors and Erg. Society Europe*, 2020.
[24] H. R. Pelikan and M. F. Jung, "Designing robot sound-in-interaction: The case of autonomous public transport shuttle buses," in *Proc. Int'l Conf. on Human-Robot Interaction (HRI)*, 2023.
[25] M. Marge, C. Bonial, B. Byrne, T. Cassidy, A. W. Evans, S. G. Hill, and C. Voss, "Applying the wizard-of-oz technique to multimodal human-robot dialogue," in *Proc. RO-MAN*, 2016.
[26] D. Strazdas, J. Hintz, A.-M. Felßberg, and A. Al-Hamadi, "Robots and wizards: An investigation into natural human–robot interaction," *IEEE Access*, vol. 8, 2020.
[27] B. Pfleging, N. Seyffer, Z. Beek, and S. Zug, "A wizard-of-oz setup to assess pedestrian behavior and teach delivery robots on sidewalk encounters," in *Robots for Public and Social Spaces - Design for Intent Communication, Collaboration and Acceptance*, 2023.
[28] A. Vega Vega, K. D. Ramírez Benavides, L. A. Guerrero Blanco, and G. López Herrera, "Evaluating the nao robot in the role of personal assistant: The effect of gender in robot performance evaluation," in *Proc. Int'l Conf. on Ubiquitous Computing and Ambient Intel.*, 2019.
[29] J. K. Westlund, C. Breazeal, and A. Story, "Deception, secrets, children, and robots: What's acceptable," in *HRI Workshop on The Emerging Policy and Ethics of Human-Robot Interaction*, 2015.
[30] L. Riek and D. Howard, "A code of ethics for the human-robot interaction profession," *Proceedings of we robot*, 2014.
[31] D. J. Rea, D. Geiskkovitch, and J. E. Young, "Wizard of awwws: Exploring psychological impact on the researchers in social hri experiments," in *Comp. Int'l Conf. on Human-Robot Interaction*, 2017.
[32] P. Saulnier, E. Sharlin, and S. Greenberg, "Exploring interruption in hri using wizard of oz," in *Int'l Conf. Human-Robot Interaction*, 2010.
[33] E. Mitchell and L. Mamykina, "From the curtain to kansas: Conducting wizard-of-oz studies in the wild," in *CHI Ext. Abstracts*, 2021.
[34] C. Clabaugh and M. Matarić, "Escaping oz: Autonomy in socially assistive robotics," *Ann. Rev. of Control, Robotics, and Aut. Sys.*, 2019.
[35] K. Malterud, V. D. Siersma, and A. D. Guassora, "Sample size in qualitative interview studies: guided by information power," *Qualitative health research*, vol. 26, no. 13, pp. 1753–1760, 2016.
[36] K. Charmaz, *Constructing grounded theory*. sage, 2014.
[37] A. Shick, "Romibo robot project: An open-source effort to develop a low-cost sensory adaptable robot for special needs therapy and education," in *ACM SIGGRAPH 2013 Studio Talks*, 2013.
[38] PEERbots. (2021) Peerbots. [Online]. Available: https://peerbots.org
[39] S. Elbeleidy, T. Mott, and T. Williams, "Practical, ethical, and overlooked: Teleoperated socially assistive robots in the quest for autonomy," in *Comp. Int'l Conf. on Human-Robot Interaction (alt.HRI)*, 2022.
[40] S. Elbeleidy, D. Rosen, D. Liu, A. Shick, and T. Williams, "Analyzing teleoperation interface usage of robots in therapy for children with autism," in *Proc. Interaction Design and Children*, 2021.
[41] S. Elbeleidy, A. Jadhav, D. Liu, and T. Williams, "Robot teleoperation interfaces for customized therapy for autistic children," in *Proc. Int'l Conf. on Human-Robot Interaction (HRI)*, 2022.

APPENDIX

1) What kind of robots do you use/wish to use in the future?
2) Are those physically embodied or simulated?
3) What aspects of the robot/s do you need control?
4) In your work and expected future work, how many physical robot bodies are involved in research studies?
5) How many robot personas/identities are involved in research studies?
6) Are those personas only associated with a single robot body or does the association between robot bodies and personas change?
7) What are the situations in which you would need to control more than one robot at a time?
8) When conducting a human-robot interaction study, how many robots would a single person control at a time?
9) When conducting a human-robot interaction study, what devices and programs do you use to control the robot/s?
10) What do you like about using these devices and programs?
11) What does the preparation process look like for connecting to a robot?
12) When using a robot control interface, what about an interface matters to you most?
13) Is there a particular performance quality that is/would be important in your work?
14) What are particular challenges or robot failures you have come across when controlling robots?