Determining appropriate first contact distance: trade-offs in human-robot interaction experiment design

Aaron Cass, Eric Rose, Kristina Striegnitz and Nick Webb[1]

Abstract— Robots are increasingly working in close proximity to and in collaboration with people. Yet much is still unknown about the best behaviors for robots to use in interactions with humans. For example, in approaching someone for help, is there an optimal distance from a candidate person at which the robot should attempt to initiate contact? What distance is neither so far away as to risk not getting the person's attention nor so close that the person has already committed to passing the robot?

Experiments that attempt to answer such questions must balance tensions. In this paper, we argue for relatively low-cost, rapid turnaround experiments achieved by carrying out experiments "in the wild" in partially controlled ways using Wizard of Oz robot control with on-board sensing. We present the trade-offs of such experiments and illustrate the approach with a set of experiments we are carrying out to determine an appropriate distance for initial contact.

I. INTRODUCTION

We envision a future in which robots mingle with humans in city squares, train stations, malls, or popular meeting spots on college campuses. These robots will perform a variety of tasks such as cleaning, leading guided tours, delivering items or providing help and information. To safely and successfully share the space with humans, these robots need to be aware of the people around them and must be able to interact with them in socially appropriate ways [1]. They need to be able to predict the humans' actions and to behave in ways that are natural and predictable to humans. This requires knowledge about how humans move through space and the ability to adapt the robot's own movements to the movement patterns of the people around it. The robot also needs to be able to communicate with people in a way that is natural and easy to understand for humans without requiring any training of or by the humans.

These robots may have to approach people to ask for their help or cooperation. For example, a robot might need to get directions to a specific person for whom it has a message [2], to ask people to make way as a tour is passing through [3], or to get assistance calling the elevator to deliver a package to another floor [4]. In all of these examples, the robots will need to approach nearby people and interrupt those people's ongoing activities to ask them for help. In other words, the robots need to draw people into ad-hoc collaborations.

Approaching a person for help involves three main sub-tasks: selecting who to ask, initiating contact and actually approaching that person, and engaging the person in a dialog using verbal and non-verbal communicative behaviors. Before we can develop mechanisms for selecting, approaching, and engaging, there are many research questions to answer. For example, what characteristics of a person or the person's behavior make a person an ideal candidate for interaction? Presumably, a person in a hurry is not likely to stop to offer assistance, but what sensor reading corresponds to "in a hurry"? Equally, we would like to choose a person who is not so far away that we cannot get their attention, and not so close that they have already committed to passing the robot. What then is the *appropriate* range of distances at which to attempt interaction? Our current research focuses on determining strategies for approaching people that are most successful in getting the robot the necessary help.

The field of social robotics is new and the answers to many of these questions are still unknown. Experiments are necessary to inform the design of social robots and to evaluate system prototypes. In this paper, we describe the design of experiments that we are currently running to answer some of these questions, and we discuss some of our experiment design choices. In particular, we have decided to run studies in the wild rather than in a laboratory. Furthermore, we use a Wizard of Oz design, where the participants think that the robot is autonomous, while it is, in reality, controlled by a human experimenter. Finally, we use low-cost sensing and, in particular, avoid expensive environmental sensors. These decisions entail trade-offs, but we argue that, overall, they give us results that are more likely to carry over to real deployed situations. In addition, they make it possible to set up studies with a quick turnaround time, enabling an iterative approach to the design and development of the social robot [5], and they allow us to flexibly change the location of our studies.

After an overview of studies addressing similar research questions from the literature in Section II, we will discuss the trade-offs entailed by the choices we made in designing our experiments in Section III. Then we will illustrate our approach by describing a sequence of studies designed to determine what is the best distance for a robot to first initiate contact with a human passer-by in Section IV. Section V summarizes our conclusions.

II. RELATED WORK

The studies we describe in this paper are part of an iterative process intended to discover the best method for approaching a human to interact with. Having selected a person to interact with (by methods not described here) we could simply move in their direction, closing the distance

[1]Department of Computer Science, Union College, Schenectady, NY, USA. {cassa,rosee,striegnk,webbn}@union.edu

between the person and the robot. However prior work suggests that there are several factors in approach behavior that may impact the success of interaction attempts (e.g., the sequence of behaviors that the robot and the human use when they first initiate contact and approach each other [6], [7], the direction of approach [8], [9] and the distance between the robot and the subject while interacting [10], [11], [12]).

Heenan et al. [7] model the greeting behavior of a robot on the detailed description by Kendon and Ferber [6] of the behaviors people use when they first see their interaction partner, establish contact, exchange a first salutation, approach the other person, greet him or her and finally open the conversation. Heenan et al.'s evaluation of their robot is preliminary in that they only report on informal observations and don't systematically compare different instantiations of their overall model. In the experiments described later in this paper we are trying to narrow down what the optimal distance is at which the robot should first call attention to itself and signal its desire to interact with the human, a question which has been left unexplored by previous work.

In the work of Satake et al. [8], subjects were selected for interaction, and approaches made via the shortest path. However the authors noted a number of failed attempts, in part because the subjects failed to notice that the robot was attempting to engage with them, especially when the robot approached from the side. To solve this, they implemented a frontal approach strategy where the robot detected a subject and maneuvered to be in front of them, a goal which requires significant long-range planning and trajectory computation so as to be in the correct place, with the correct orientation (i.e. facing the subject) before engaging. This method increased the number of successful interactions, but relies on the use of substantial environmental sensing. In their work, Satake et al. made use of six LIDAR sensors placed statically in a shopping environment.

Prior work in a more enclosed, living room scenario by Sisbot et al. [9] showed that subjects preferred the robot to approach from the side, as opposed to frontal approaches. In their scenario subjects were seated in a chair in the center of the room watching TV and the robot would approach with a remote control. 80% of the subjects rated the frontal approach as the least preferred option (the most preferred was from the right). They conjecture that this result was impacted by the specific scenario. Most subjects sat in the chair with their legs directly out in front of them, and might have been anxious about contact between the robot and their legs. This preference could potentially change if the subjects had a better sense of where the robot is "looking."

This direction of gaze effect was seen in work by Takayama and Pantofaru [10], where the "gaze" of the robot (at the feet vs. at the face) affected the acceptable approach distance recorded by human subjects. Takayama and Pantofaru hypothesized that mutual gaze, when the robot and the subject are looking at each other's faces, increases the size of the comfortable space between them (and hence the preferred distance for interaction), but although they were unable to support this hypothesis, they did find a correlation between gaze direction and ideal approach distance based on gender; when the robot was looking at the subject's face, men got closer to the robot than women.

The main focuses of the Takayama and Pantofaru study were the factors that impact ideal approach distance, both when a robot approaches a subject and when a subject approaches a robot. Their findings, along with those of Walters et al. [11], suggest that subjects treat robots similarly to how they treat human "social partners," with approach distances of 0.5m to 1.25m that align with those of human-to-human interactions (see Hall [13]). Both robot studies examined features of the subjects that impact comfortable approach distance, such as prior experience with robots (which reduced approach distance), or with pets (also resulting in reduced comfortable approach distance). Prior experience with robots was determined by a questionnaire, measuring greater than or less than one years experience with robots. Walters et al. [14] subsequently performed an experiment that examined the longitudinal effect of sustained interaction with a robot, with a small number of participants (initially 7, dropping to 4 by the end of the process), in a controlled environment. One issue with all of these studies is that they were conducted under laboratory, or highly controlled, conditions.

We hypothesize that these results may be different when conducted in the wild, with subjects who do not know at the time of the study that they are being monitored. There are also confounding issues in successful approaches. Most prior work examines comfortable approach distance between robot and human; that is, how close to each other can they get, before the human feels uncomfortable. Our interest lies in the most appropriate contact distance, that is, how far away from a subject should a robot attempt interaction, or inversely, how close is too close for a robot to attempt an interaction with a subject who is not expecting to interact with a robot. The appropriate contact distance might be one in which the human is not comfortable, although we would not be surprised to find through experimentation that the most appropriate distance for first contact correlates with Hall's *public* or *social* distances [13]. For this research, the approach speed, distance and angle are the physical factors, but these ultimately have to be coordinated with appropriate interaction attempts, through voice, gesture, use of display and other efforts.

III. EXPERIMENTAL DESIGN DISCUSSION

We now discuss some tensions that experiments designed to study human robot interaction in public spaces have to balance:

- **Level of experimental control.** An experiment will be undertaken in either a controlled laboratory setting or "in the wild", or in an environment with some level of experimental control between these two extremes. The trade-off has an effect on the reliability and generalizability of the results.
- **Robot control mechanism.** The robots used in an experiment need to have their behavior controlled. The mechanism of control can require more or less human

intervention (i.e. the control can be less or more autonomous). The trade-off has an effect on the cost and timing of experiments.

In the rest of this section, we discuss these trade-offs in more detail.

A. Laboratory vs. "In the wild" Experiments

In contrast to laboratory experiments, where one can control many aspects of the participants and the environment, when studying human-robot interaction "in the wild" we do not have full control over who participates or what the environment looks like. For example, we don't know how familiar the participants are with robots and technology in general, there may be other people in the environment, other ongoing activities may sometimes distract from the robot, and the participants may be focused on where they are going rather than the interaction with the robot. This complicates the interpretation of data collected in the wild. On the other hand, the circumstances in laboratory evaluations may be too idealistic and may lead to results that cannot be replicated when the system is deployed in the "real world". For example, in the world of spoken dialogue systems, it has long been known that evaluation in the lab does not give an accurate forecast of the behavior of a system once it has been deployed. Van Haaren et al. [15] discuss one such problem with the Dutch rail assistant telephone agent, which had a task completion percentage of 94% when evaluated in the lab by volunteers, but a task completion of 68% when actively deployed in the wild. The explanation from the authors makes it clear that while laboratory controlled experiments are useful, the behaviors of participants in real life are significantly more varied and harder to predict.

In addition, for some research questions, setting up a laboratory experiment may be quite difficult. In the example studies discussed in this paper we are trying to determine what is the best distance for first contact when a social robot wants to ask an unsuspecting human for help. We set up a robot in different positions in well-trafficked areas on a college campus and observe what happens when people come across the robot unexpectedly. To achieve a similar level of surprise in the laboratory, we would have to employ a deception scheme where we recruit participants to come to the lab for an ostensible study on some unrelated topic and then make sure that they come across the robot as they enter the lab. While this is harder to design, set up and carry out, it is still not clear that participants would react to the robot in the same way when they come across it in a research lab as when they encounter it in the course of their normal daily activity.

Finally, laboratory studies tend to be more costly and time consuming, so that studies in the wild will often be able to produce more data. In the studies described later in this paper, we set up a robot in a thoroughfare on a college campus, which allowed us to observe several interactions during the same time frame that in the lab we would have required to capture one interaction. This additional volume of data also makes it possible to deal with some of the noise introduced by not being able to control every aspect of the experiment. For example, in a separate study [16] that collected data over the Internet as well as in a laboratory experiment to evaluate an interactive instruction giving system, we found that while the data from the Internet based study is more noisy (with, for example, users canceling after only partially completing the task, some users leaving many questionnaire items blank, and a skewed gender distribution), the results of the Internet based study and the lab experiment are consistent, and we were able to collect so much more data over the Internet, that a more in-depth analysis was possible. It would not have been feasible for us to collect a similar amount of data in the lab because of time and budget constraints.

B. Trade-offs in human vs. autonomous control

Another tension that any experimental design must address relates to the level of autonomous control needed in the experiments. At one end of the spectrum is full autonomous control, which has the primary benefit of fidelity. Because the robot behaves autonomously, and thus responds to sensor inputs in the same way every time, when people respond to the robot's actions they are responding to the same actions the robot would take in a real-life scenario. If the ultimate goal is to determine robot behaviors that elicit desirable human responses, a fully-autonomous robot ensures that the robot behaviors are repeatable.

However, developing an autonomous robot behavior requires significant time and expense, so waiting until algorithmic control can be developed before experiments can begin means delayed validation of the approach. Worse, we don't know the best robot behaviors until *after* experiments are completed. For example, before we've carried out experiments, we don't know the optimal distance at which the robot should first initiate contact with the human and signal its intention to interact. We could attempt to develop an autonomous system in which this *contact distance* is a configurable parameter, but that would likely be more difficult than developing such a system capable of a much narrower range of contact distances. Furthermore, different contact distances might require very different sensor technology — for example, close contact distances might be easy to manage with on-board sensors, while longer-range contact distances might call for more expensive on-board sensors or even environmental sensors. Because the appropriate contact distance we can autonomously manage will depend on the sensor regime in place, we risk making a system that is more complex, and expensive, than needed.

These problems are ameliorated if we can carry out experiments with a robot *before* autonomous control is implemented. Borrowing from HCI approaches, we propose Wizard of Oz (WoZ) studies, in which the robot is portrayed as autonomous while actually being controlled by an operator out of view of the subjects. Thus, subjects in the experiment have the experience of interacting with an autonomous robot even though the sensor regime and algorithms are not yet in place to have the robot behave autonomously. Riek [17] gives an excellent overview of the use of WoZ studies in

Human-Robot Interaction, including challenges faced when performing these kind of studies with robots. In line with the suggestions contained in this paper, we consider our WoZ studies part of an iterative process in the development of autonomous robot control. Specifically, we use these methods as a way to guide the direction of development of the robot. This presents its own challenges, including making sure that we simulate robot behavior in a manner that can ultimately be automated.

In order for us to trust the data we get about human responses to robot behavior, the robots must *seem* real to the subjects in our experiments. If a robot displays much more intelligence than can later be implemented in an autonomous control system, then human response to it might differ markedly from what can later be deployed. To accomplish the goal of producing behaviors that seem real to experimental subjects, we script the behaviors the robot can perform and then restrict the human wizard to follow these scripted behaviors. This can often be enforced by convention, but in some cases it is easy enough to enforce constraints on the robot behavior more strictly by providing an interface for the wizard that only allows for certain behaviors or by using autonomous behaviors built into the robot for certain simple subtasks.

In this way, we are clear about both our expectations of the robot, and of the people who interact with them. What is novel in our suggested experiment is that the humans in the study are not expecting to encounter a robot. There is no predefined scenario for the human, instead the robot has a task to complete with wizard assistance, and attempts to approach unsuspecting subjects. The completion of the task (say, answering a simple question) is one facet of the evaluation behavior, with analysis of the human reaction to the attempted approach the key factor in this experiment.

IV. CONTACT DISTANCE EXPERIMENTS

To illustrate our approach to social robotics experiment design, addressing the trade-offs outlined above, we will describe a set of experiments we have designed and carried out to help determine the appropriate range of distances at which a robot's attempts to initiate an interaction with a human are most likely to be successful.

We started with preliminary *exploratory* studies in which we were trying to determine how people react to an autonomous robot in a busy building on our college campus. Because we were trying to determine robot behaviors that elicit different human responses, we did not yet know what robot behaviors to try. Therefore, the Wizard of Oz approach seemed like a natural fit — we could undertake the experiments *before* expending much time and expense to develop autonomous control, and *before* we knew what behaviors an autonomous controller must support.

The main purpose of these initial experiments was to figure out who the ideal person(s) would be to approach, and how to approach them. Because these studies were exploratory, the human robot controller (*the wizard*) was not trying to follow any pattern or accomplish any goal with the robot.

From a hidden spot, the controller would drive the robot around and have it approach random people, recording their reactions with on-board cameras. Because the system was under human control, and the human controller could see the interactions, we did not need more expensive sensors — we only used on-board cameras because it made off-line analysis easy.

Our studies were conducted using the TurtleBot platform,[1] comprised of a mobile base with an attached depth sensor such as a Microsoft Kinect. The TurtleBot is supported under ROS - the open source Robot Operating System.[2] The experimental set up of our TurtleBot can be seen in Fig. 1a, with three cameras to capture interactions with passers-by. We know that future interaction experiments require a human scale robot, and we have an experimental platform which we will use when initial hypotheses have been validated. LINDSEE (Learning Interactive Navigator and Developmental Social Engagement Engine), designed and built by our students, can be seen in Fig. 1b. We used the TurtleBot for these initial experiments as the full capabilities of LINDSEE were not required to validate our initial hypotheses, and we can continue development on LINDSEE as these experiments progress. The white mast just visible on the TurtleBot in Fig. 1a extends the height of the robot to the same 5' height as LINDSEE. While the TurtleBot is different from LINDSEE in important ways, in the experiments we describe here, subjects interacted with the TurtleBot in ways that indicated they saw it as an independent social agent (e.g. waving at the robot). Our experiments are designed so that the robots we intend to use in this environment move and interact in the same way, and all robots use ROS, making the software we develop transferable between them.

From the pilot studies described above, a pattern emerged. When approaching a person, the initial distance between the person and the robot played a role in whether or not the person would react with some form of communication indicating that the robot had the subject's attention (such as waving, pointing to the robot, or stopping and staring at the robot). It seemed that people tended to pay more attention to the robot when it was not too far away nor too close. This led us to hypothesize that there is a bounded range of appropriate distances for the initial contact between the robot and the candidate.

The question was then, what exactly is the appropriate range of distances? Can we bound the range to a small range of distances? Again, we planned further studies to try to answer this question. But first, we started with a very inexpensive *validation study* aimed at telling whether the hypothesis was worth pursuing. The purpose of this experiment was to confirm if there was a distance that was too close — a contact distance that would result in the person being less likely to communicate with the robot. We reasoned that if our hypothesis of a range of appropriate distances was valid, then there must be a contact distance that is

---

[1]http://www.turtlebot.com/
[2]http://www.ros.org

(a) Turtlebot

(b) LINDSEE

Fig. 1: Our experimental robot platforms

too close. Performing a time-consuming study to narrow the range would be a waste of resources if the hypothesis wasn't valid at all.

The human controller in this second experiment stationed the robot inside, but near to the entrance of, a building on campus. The robot was placed in such a way that people entering the building from outside could not see the robot until they had already opened the door. In this way, we know that the distance the robot is from the door is the same as the distance between the human subject and the robot when the subject first encounters the robot. With this setup, once a person entered the building, the wizard would drive the robot towards the person.

More often than not, the person entering would actively go out of their way to avoid the robot, as if it were some obstruction on the floor. One person was visibly scared by the robot. This anecdotal observation supports our hypothesis that there is a contact distance that is too close for interaction; the person encountered the robot very suddenly, before any action on the part of the robot (or the controller in this case) could impact the interaction.

With bolstered confidence in the validity of our hypothesis, we set out to design an experiment to narrow the range of distances. This experiment needs to be more controlled than the first exploratory study, but as argued above, we avoid strictly-controlled laboratory experiments so that human responses are natural. Thus, we have a *semi-controlled semi-wild experiment*. What we have planned is another WoZ style experiment. In this setup, the robot will be placed at specific distances inside the entrance of a building on campus. The robot again will have on-board cameras to record the candidates' reactions to the robot. As in the previously-described validation study, the potential candidates, the people entering from said entrance, will not be able to see the robot until they have entered the building. Once the candidate has entered the building the robot will start its approach. We will consider the approach a success if the candidate tries to interact with the robot. For different runs of the experiment, we will vary the initial distance from the door and determine the success rate from each distance, thus helping us narrow the range of appropriate contact distances.

V. CONCLUSIONS

The field of social robotics is relatively young and there is much that is still unknown about the best ways to build robots that must interact appropriately with humans. In this paper, we have discussed one set of research questions related to appropriate ways of selecting and approaching people for assistance. While much is known about how humans approach each other, it is not clear how well those lessons translate to human-robot interactions.

Because such fundamental research questions are still open in the field, it seems to us that there is opportunity for much progress to be made with relatively non-resource-intensive approaches, thus enabling contributions from groups outside traditional robotics. In this paper, we have presented one possible approach to such inexpensive studies. While doing experiments "in the wild" does introduce problems of experiment control, and doing strictly-controlled experiments in laboratory settings introduces issues of generalizability, we can address these concerns with an approach between the two extremes. And if we perform the studies with inexpensive hardware, and under human instead of autonomous control, we can get experiments up and running more quickly than would be required by a more traditional approach of building an autonomous robot and testing it. Further, if we adhere to experimental and reporting guidelines that ensure the rigor and replicability of our work, then significant advances can be made using an iterative approach to experimentation and design.

Much can be learned, of course, from experiments performed using autonomous robots. And much can be learned from field studies and laboratory studies alike. In this paper, we argue for an approach that we think complements those, while recognizing that there are trade-offs. We have made those trade-offs in part because we believe that they constitute the best experimental set-up for our research questions and in part because of resource constraints. Others with similar constraints, or others interested in similar questions, can make similar trade-offs.

REFERENCES

[1] C. L. Breazeal, *Designing Sociable Robots*. MIT press, 2002.

[2] M. P. Michalowski, S. Šabanović, C. DiSalvo, D. Busquets, L. M. Hiatt, N. A. Melchior, and R. Simmons, "Socially distributed perception: GRACE plays social tag at AAAI 2005," *Autonomous Robots*, vol. 22, no. 4, pp. 385–397, 2007.

[3] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "A Larger Audience, Please!: Encouraging People to Listen to a Guide Robot," in *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, ser. HRI '10. Piscataway, NJ, USA: IEEE Press, 2010, pp. 31–38. [Online]. Available: http://portal.acm.org/citation.cfm?id=1734469

[4] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss, "Robots asking for directions: The willingness of passers-by to support robots," in *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*. IEEE Press, 2010, pp. 23–30.

[5] S. Šabanović, S. Reeder, and B. Kechavarzi, "Designing robots in the wild: In situ prototype evaluation for a break management robot," *Journal of Human-Robot Interaction*, vol. 3, pp. 70–88, 2014.

[6] A. Kendon and A. Ferber, "A description of some human greetings," in *Comparative Ecology and Behaviour of Primates: Proceedings of a Conference held at the Zoological Society London*, R. P. Mochael and J. H. Crook, Eds. Academic Press, 1973.

[7] B. Heenan, S. Greenberg, S. Aghel-Manesh, and E. Sharlin, "Designing social greetings in human robot interaction," in *Proceedings of the Conference on Designing Interactive Systems*, 2014, pp. 855–864.

[8] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?-strategies for social robots to initiate interaction," in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*. IEEE, Mar. 2009, pp. 109–116.

[9] E. A. Sisbot, R. Alami, T. Simeon, K. Dautenhahn, M. Walters, and S. Woods, "Navigation in the presence of humans," in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*. IEEE, Dec. 2005, pp. 181–188.

[10] L. Takayama and C. Pantofaru, "Influences on proxemic behaviors in human-robot interaction," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 5495–5502.

[11] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, Aug. 2005, pp. 347–352.

[12] R. Mead and M. J. Matarić, "Perceptual models of human-robot proxemics," in *Proceedings of the 2014 International Symposium on Experimental Robotics (ISER 2014)*, June 2014, marrakech/Essaouira, Morocco.

[13] E. T. Hall, *The Hidden Dimension*. Doubleday, 1966.

[14] M. L. Walters, M. A. Oskoei, D. S. Syrdal, and K. Dautenhahn, "A long-term human-robot proxemic study," in *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication: RO-MAN 2011*, 2011, pp. 137–142.

[15] L. van Haaren, M. Blasband, M. Gerritsen, and M. van Schijndel, "Evaluating quality of spoken dialogue systems: Comparing a technology-focused and a user-focused approach," in *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC)*, 1998, pp. 655–660.

[16] A. Koller, K. Striegnitz, D. Byron, J. Cassell, R. Dale, J. Moore, and J. Oberlander, "The first challenge on generating instructions in virtual environments," in *Empirical Methods in Natural Language Generation*, ser. LNCS, E. Krahmer and M. Theune, Eds. Springer, 2010, vol. 5980.

[17] L. D. Riek, "Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines," *Journal of Human-Robot Interaction*, vol. 1, pp. 119–136, 2012.