

# Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines

Laurel D. Riek  
University of Notre Dame

---

Many researchers use Wizard of Oz (WoZ) as an experimental technique, but there are methodological concerns over its use, and no comprehensive criteria on how to best employ it. We systematically review 54 WoZ experiments published in the primary HRI publication venues from 2001 - 2011. Using criteria proposed by Fraser and Gilbert (1991), Green et al. (2004), Steinfeld et al. (2009), and Kelley (1984), we analyzed how researchers conducted HRI WoZ experiments. Researchers mainly used WoZ for verbal (72.2%) and non-verbal (48.1%) processing. Most constrained wizard production (90.7%), but few constrained wizard recognition (11%). Few reported measuring wizard error (3.7%), and few reported pre-experiment wizard training (5.4%). Few reported using WoZ in an iterative manner (24.1%). Based on these results we propose new reporting guidelines to aid future research.

Keywords: Wizard of Oz, human-robot interaction, systematic review, reporting guidelines, methodology

---

## 1. Introduction

As robots move more toward being “in the loop” and having increased autonomy in their interactions with humans, it is important that Human-Robot Interaction (HRI) practitioners and researchers are employing scientifically rigorous practices during the robot design lifecycle.

One commonly employed technique in the HRI researchers’ toolkit is the Wizard-of-Oz (WoZ) technique (Kelley, 1984). WoZ refers to a person (usually the experimenter, or a confederate) remotely operating a robot, controlling any of a number of things, such as its movement, navigation, speech, gestures, etc. WoZ may involve any amount of control along the autonomy spectrum, from fully autonomous to fully tele-operated, as well as mixed initiative interaction.

Researchers who employ WoZ argue that because robots are not sufficiently advanced to interact autonomously with people in socially appropriate or physically safe ways, this kind of puppeteering allows participants to envision what future interaction could be like. They also might use it to test out early aspects of their design that have yet to be fully implemented, as part of an iterative design process.

However, some researchers have raised methodological concerns regarding the use of this technique. For social interaction, Weiss (2010) suggests that a WoZ controlled robot is serving more as a proxy for a human and less as an independent entity. Thus, it is not really human-robot interaction so much as human-human interaction via a robot.

---

Authors retain copyright and grant the Journal of Human-Robot Interaction right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work’s authorship and initial publication in this journal.



*Figure 1.* “Wizard” clinicians controlling a robotic patient mannequin in a medical simulation scenario.

Others have raised ethical concerns about the use of WoZ and social deception. Fraser and Gilbert (1991) discusses participant gullibility, and subsequent embarrassment, over finding out they have been deceived. Fraser and Gilbert (1991) also discuss ethical conundrums faced by the experimenter in terms of how to mitigate the necessity of deceiving a participant to keep the simulation realistic against the act of deception.

Riek and Watson (2010) and Miller (2010) also suggest ethical problems when participants can not tell whom or what they are interacting with - a particular human, a human masquerading as another human, or a machine. (Miller refers to these as “Turing Deceptions”). In the UK, the funding agency EPSRC drafted ethical rules for robots, and went so as far to state that users should always be able to “lift the curtain,” since robot intelligence is, currently, an illusion (EPSRC, 2011).

From an engineering perspective, Breazeal et al. (2005) suggest that relying on WoZ as an experimental technique can make it all the more difficult to build robots capable of successfully mitigating errors on their own in the future (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005).

One way to mitigate these methodological, ethical, and engineering concerns is to design experiments that employ WoZ in a rigorous and repeatable manner, allowing for a smoother transition to a more autonomous and capable system in the future.

In the field of HRI we traditionally have not had explicit criteria to guide us in this task, but the Natural Language Processing (NLP) community does. According to Fraser and Gilbert (1991), to conduct a valid WoZ simulation the following requirements must be met:

1. “It must be possible to simulate the future system, given human limitations;”
2. “It must be possible to specify the future system’s behavior; ”
3. “It must be possible to make the simulation convincing.”

Fraser and Gilbert (1991) discuss controlling for various aspects of the Wizard’s behavior, such as Wizard recognition variables (i.e., is the Wizard perfect? How is error controlled for?), Wizard production variables (i.e., how fast should the Wizard respond?), and Wizard training (i.e., giving

the Wizard enough training time to be as similar as possible to the envisioned autonomous system). While Fraser and Gilbert (1991) suggested these requirements in the context of non-embodied speech and dialogue systems, they are very applicable HRI WoZ systems as well.

Another important consideration for WoZ experiments is to employ the use of scenarios to place additional constraints on the study. Dahlbäck et al. (1993) and Fraser and Gilbert (1991) stress the importance of scenario constraints for NLP WoZ experiments, as do Green, Hüttenrauch, and Eklundh (2004) for HRI. This allows researchers to provide participants with “a task to solve whose solution requires the use of the system, but where there does not exist one single correct answer, and/or where there is more than one way to reach the goal” (Dahlbäck, Jönsson, & Ahrenberg, 1993).

Thus, Green et al. (2004) suggest the following method for WoZ HRI scenario construction:

**User instruction** gives information to the user to help them perform the intended task. (i.e., “What should the user do?”)

**Behavior hypotheses** reflects the designer’s beliefs about users actions within the scenario. (i.e., “What will the user do?”)

**Robot behavior** specifies what the robot will do, for any level of autonomy. (i.e., “What should the robot do?”)

Steinfeld, Jenkins, and Scassellati (2009) argue that it is important for WoZ studies to also take into account how the environment can affect both the robot and the human, how the level of autonomy of the robot can affect how humans control it, and, in some cases, how to model, simulate and otherwise account for the actions of the Wizard.

Thus, one may need to design studies that employ one or more human or robot centric experimental techniques throughout the design lifecycle, including not only *Wizard of Oz*, but also *Wizard with Oz* (human-centric method that uses real technology in a simulated environment), *Wizard and Oz* (human-centric method that uses real technology in a real environment), *Oz with Wizard* (robot-centric method that includes humans but doesn’t measure them), *Oz of Wizard* (robot-centric method where humans are either simulated or minimally involved), and *Wizard nor Oz* (all aspects of the system are simulated).

Finally, returning to the original article by Kelley (1984) who first introduced the WoZ technique to the usability community, iteration is a key component of valid design. A single WoZ simulation was not intended to be the end point of a research question, but part of an iterative process in which the Wizard is gradually phased out of the communication loop and autonomy is phased in (Kelley, 1984).

Thus, it is with the WoZ methodological criteria proposed by Fraser and Gilbert (1991), Green et al. (2004), Steinfeld et al. (2009), and Kelley (1984) that we perform a systematic review of the HRI literature to synthesize existing evidence of how researchers are currently designing and conducting WoZ studies. A systematic review “involves the application of scientific strategies, in ways that limit bias, to the assembly, critical appraisal, and synthesis of all relevant studies that address a specific... question” (Cook, Mulrow, & Haynes, 1997).

Section 2 of this paper describes our review protocol and methodology, Section 3 describes our results, and Section 4 discusses the implications of these results for the HRI community and proposes new reporting guidelines to guide future research with WoZ.

## 2. Review Protocol

We searched the ACM Digital Library (ACM, 2012) and IEEEExplore (IEEE, 2012), as these include the major peer-reviewed HRI and general robotics publication venues. While there are several HRI

Table 1: Literature review search strategy.

Search Terms	Source Publications
("wizard of oz" OR WoZ) AND "human robot interaction" AND (experiment OR "user study" OR pilot)	ACM International Conference on Human Robot Interaction (HRI)
	IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
	IEEE International Conference on Robotics and Automation (ICRA)
	IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)
	IEEE Transactions on Systems, Man, and Cybernetics (SMC)
	IEEE Transactions on Robotics (TRO)

publication venues not indexed by ACM or IEEE, for example *Interaction Studies* and *The Journal of Social Robotics*, it was necessary to constrain the search space, so they were not included in this review. The publications we included and the search terms used are detailed in Table 1.

We included papers in this systematic review that (1) had been published in a peer-reviewed conference or journal, (2) included at least one physically embodied robot that interacted with at least one human participant, and (3) described an experiment or user study that included five or more participants.

This review excluded unpublished or non-peer-reviewed manuscripts, book chapters, posters, talks, demonstrations, exhibits, videos, and abstracts. We also excluded "lightly peer reviewed" publications, such as HRI Late Breaking Reports and CHI Extended Abstracts.

Any papers published before 2001 were also excluded from the search. Since robotics as a discipline advances so rapidly, studies published before 2001 may not reflect the state of the art in terms of robot capability.

Papers that solely described simulated platforms or virtual avatars were excluded from this review, as well as any papers that only referenced or alluded to a WoZ study or experiment but did not actually describe one. Finally, if an author reported the exact same study or experiment in two or more different papers, only one paper was included and the others were excluded.

### 3. Results

#### 3.1 General Paper Demographics

The original search returned 177 results. After removing duplicates across the two databases and applying the exclusion criteria, 54 papers were included in the analysis. The included papers were published from 2003 - 2011, the mode being 2009.

Figure 2 depicts the major publication venues. IEEE RO-MAN and ACM HRI were the major venues the included studies appeared in at 38% and 32% respectively. The other major IEEE

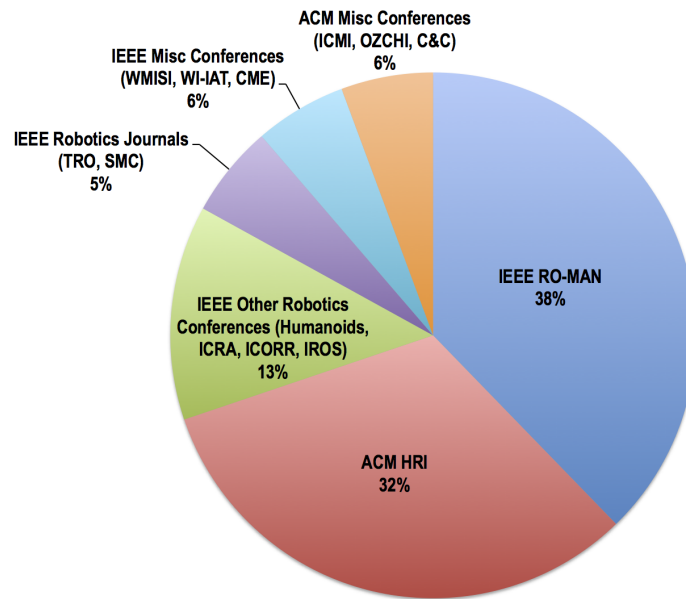


Figure 2. Pie chart showing the distribution of included WoZ articles in this review. Seventy percent appeared in proceedings of the ACM International Conference on Human-Robot Interaction (HRI) and IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).

Robotics Conferences (Humanoids, ICRA, IROS, and ICORR) held 13% of the publications. Only 5% of the included studies appeared in IEEE robotics journals (T-RO and SMC Part II). The remaining 12% of publications appeared in other ACM and IEEE conferences, such as OZCHI, Creativity and Cognition, etc.

Figure 3 depicts the types of robots used across the 54 papers. 4 papers used androids (e.g., Repilee), 18 papers used humanoids (e.g., NAO), 3 papers used both humanoids and mechanical-looking robots (e.g., Robovie and a Cart robot), 20 used mechanical-looking robots (e.g., Pioneer 3AT), and 9 used zoomorphic robots (e.g., Pleo).

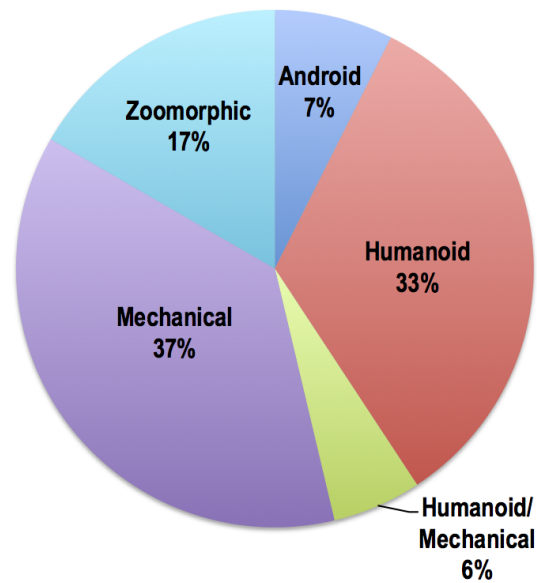
The number of participants in the studies/experiments ranged from 5 to 159; the mean was 37.96 (s.d. = 39). If a paper contained more than one experiment/study that met the inclusion criteria, the participant counts were summed.

### 3.2 Types of Wizard control

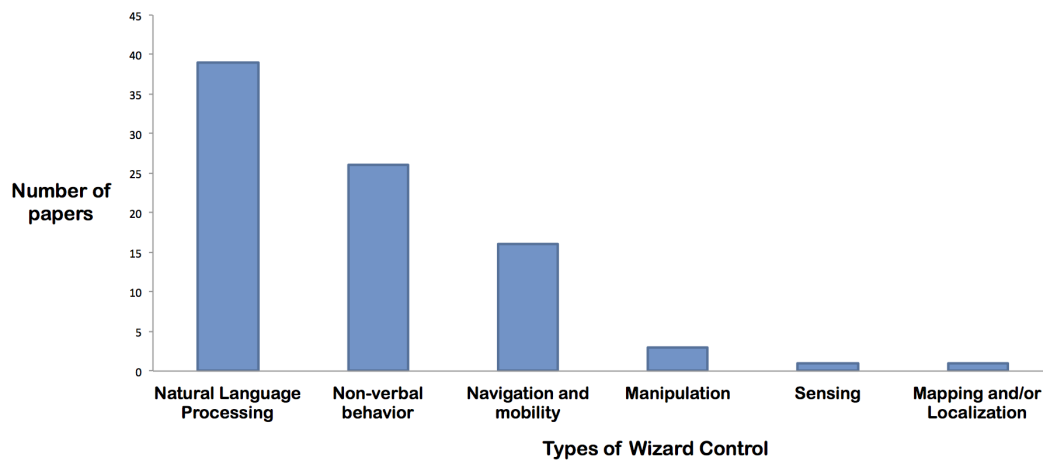
Figure 4 displays the types of Wizard control employed across the 54 papers. Some papers described using more than one type of control.

**3.2.1 Natural language processing.** In 72.2% of papers ( $n = 39$ ), authors reported using Natural Language Processing as a component controlled by a Wizard. For example, having the robot appropriately respond to things the user said to it, making utterances, etc.

**3.2.2 Nonverbal behavior.** In 48.1% of papers ( $n = 26$ ), authors reported having the Wizard make the robot generate some sort of visible utterance (i.e., non-verbal behavior). For example, making the robot nod its head, point, fix its gaze upon a user, etc.



*Figure 3.* Pie chart showing the types of robots used in experiments across the included papers. Androids were used in 7% of papers (e.g., Repilee), 33% used humanoids (e.g., NAO), 6% used both humanoids and mechanical-looking robots (e.g., Robovie and a Cart robot), 37% used mechanical-looking robots (e.g., Pioneer 3AT), and 17% used zoomorphic robots (e.g., Pleo)



*Figure 4.* Chart depicting the types of Wizard control employed in the included papers. Some papers described using more than one type of control.

**3.2.3 Navigation and mobility.** In 29.6% of papers ( $n = 16$ ), authors reported having the Wizard perform navigation and mobility tasks. For example, positioning the robot at a particular stage of the experiment.

**3.2.4 Manipulation.** In 3.7% of papers ( $n = 2$ ), authors reported having the Wizard perform manipulation tasks. Ralph and Moussa (2008) describe a task where the Wizard takes both user's commands and the robot's limitations into account when selecting commands for it; Robins et al. (2008) describe a Wizard controlling a 3-DOF drumming robot.

**3.2.5 Sensing.** In 1.9% of papers ( $n = 1$ ), authors reported having the Wizard do sensing for the robot. This sensing component was part of a semi-autonomous system Yamaoka et al. (2010) described for visual object selection.

**3.2.6 Mapping and/or localization.** In 1.9% of papers ( $n = 1$ ), authors reported having the Wizard do mapping and/or localization for the robot. Fischer (2011) described using a Wizard to localize a robotic wheelchair for a user.

### **3.3 Fraser and Gilbert (1991) Criteria**

**3.3.1 Possible to simulate the future system, given human limitations.** In 92.6% of included papers ( $n = 41$ ), authors clearly articulated that the component(s) of the system being WoZ controlled was feasible to simulate, meaning either the component could be realized in the near future, or else it was within the realm of something a human could do. For example, Okita and Sarvadevabhatla (2009) describe using a Wizard to simulate speech recognition for children interacting with ASIMO, as robust child-speech recognition is still in its infancy.

**3.3.2 Possible to specify the future system's behavior.** In 94.4% of included papers ( $n = 51$ ), authors clearly articulated that the component(s) of the system being WoZ controlled was specifiable, meaning the researchers formulated a detailed specification of the future system's behavior. Some went as far to state this explicitly in the Methodology section of their paper, for example, Burghart et al. (2007) wrote, "While the studies presented in this work have been conducted as a Wizard-of-Oz experiment, the functionality of the system has been designed to match the functionality that can be developed with a standalone system" (Burghart, Holzapfel, Haeussling, & Breuer, 2007).

**3.3.3 Constrained Wizard recognition ability.** Only 11% of included papers ( $n = 6$ ) described controlling what the Wizard could recognize and perceive in the interaction. For example, Shiomi et al. (2008) describe a semi-autonomous robot at a train station. The Wizard only intervenes in interaction when the autonomous system requests it, and furthermore, "does not listen to the interaction unless the robot asks for help so that [the wizards] could fairly well observe whether [their] semi-autonomous [system] works."

In 81.5% of included papers ( $n = 44$ ), authors did not control for what the Wizard could recognize during the interaction. Thus, the Wizard had a perfect audio and/or video feed from the experiment, the data sent to the Wizard was not explicitly delayed, etc.

**3.3.4 Constrained Wizard production ability and Wizard errors.** Although 90.7% of included papers ( $n = 49$ ) described controlling for what constraints the Wizard had while operating the robot, only 3.7% ( $n = 2$ ) of included papers clearly described the Wizard deliberately making errors when controlling the robot, and only 3.7% ( $n = 2$ ) of papers measured or discussed accidental Wizard errors. (These four papers were different). No papers described actually controlling for Wizard error, deliberate or otherwise.

One example of deliberate Wizard errors: Tanaka et al. (2007) describe an experiment where a Wizard controlled an AIBO to half the time behave randomly, mimicking a “before learning phase,” and the other half of the time behave correctly only 80% of the time, mimicking a “during-learning phase”. And one example of measuring Wizard error: Shiwa et al. (2008) describe sampling 237 experimental trials to measure Wizard error and found very few.

Only 1.9% of included papers ( $n = 1$ ) explicitly stated the authors did not control for Wizard production ability, and 7.4% of papers ( $n = 4$ ) did not provide sufficient detail to make an assessment.

**3.3.5 *Provided training to Wizards.*** Only 5.4% of included papers ( $n = 3$ ) described providing the Wizard training. Ralph and Moussa (2008) provided their Wizard *two months* of training prior to beginning their experiment, to ensure that the Wizard “did not influence each participant’s choice of language or commands used during the study.”

The remaining 94.4% of papers ( $n = 51$ ) did not provide sufficient detail about Wizard training to make this assessment. Thus, it is not clear whether such training took place.

**3.3.6 *Possible to make the simulation convincing.*** In 72.2% of included papers ( $n = 39$ ), authors clearly articulated that the component(s) of the system being WoZ controlled was convincing to participants, meaning that the illusion of interacting with an autonomous robot was maintained. For instance, a few papers described ways in which they simulated robot response lag time; Funakoshi et al. (2008) inserted random, system-level communication delays between 0 and 15 seconds to both the Wizard and the User throughout their experiment. Other researchers employed scenario constraints to immerse users in the experiment, see Section 3.4.1 for more details.

The remaining 27.8% of included papers ( $n = 15$ ) did not provide sufficient detail to make this assessment. Thus, it is not clear if their WoZ simulations were convincing to participants.

### **3.4 Green et al. (2004) Criteria**

**3.4.1 *Specified user instruction.*** In 94.4% of included papers ( $n = 51$ ), authors clearly specified what the users could do during the interaction and/or gave them a clear scenario. For example, Siino et al. (2008) gave participants a scenario where they were on a futuristic spaceship and had to command a robot to help them perform repairs to the life support system.

**3.4.2 *Specified behavior hypotheses.*** In 79.6% ( $n = 43$ ) of included papers, authors clearly specified their hypotheses for user behavior during the interaction. For example, Kim et al. (2009) test three behavioral hypotheses for users interacting with a Pleo robot in terms of their affective vocalizations.

The remaining 20.4% ( $n = 11$ ) of included papers did not clearly state their behavioral hypotheses for users, or otherwise did not provide sufficient detail to make this assessment.

**3.4.3 *Specified robot behavior.*** In 98.1% ( $n = 53$ ) of included papers, authors clearly specified the robot’s behavior during the interaction. For example, Heerink et al. (2006) described the communicative scripts they developed for their iCat robot and detailed its social expressivity, including smiling, attentive listening, and so on.

### **3.5 Kelley (1984) Criteria**

In only 24.1% ( $n = 13$ ) of included papers did authors clearly articulate that their use of WoZ was part of an iterative design process. For example, Lee et al. (2009) describe several prototype iterations as they developed Snackbot, using WoZ in early parts of their process.



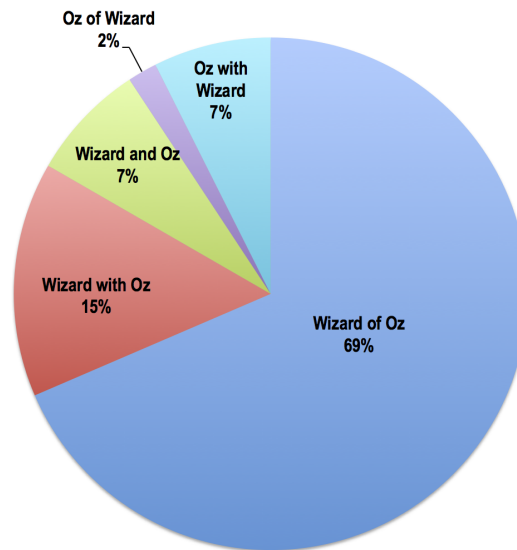


Figure 5. Pie chart depicting the types of WoZ models used across the 54 included papers.

Of the remaining 75.9% ( $n = 41$ ) of included papers authors did not provide sufficient detail to make this assessment. Thus, it is unclear if they were utilizing WoZ in an iterative way.

### 3.6 Steinfeld et al. (2009) Criteria

Figure 5 displays the types of WoZ models used across the 54 included papers. In 68.5% of included papers ( $n = 37$ ), authors described employing the *Wizard of Oz model*, wherein the robot behavior is simulated and the researcher is primarily interested in measuring how the user interacts with the robot, not measuring the robot itself. In this model, frequently realistic robot behavior “can be estimated but may not be realized until years of technology advances have occurred” (Steinfeld et al., 2009). This model might be described as “human-centric,” on one extreme of the WoZ experimental spectrum.

In 14.8% of papers ( $n = 8$ ), authors described employing the *Wizard with Oz model*, a human-centric method that uses real technology, but within simulated or controlled environments, such as a laboratory. For example, Ryokai et al. (2009) test physical and digital authoring languages for children to do storytelling on a Pleo robot, and ran their study in a controlled playspace in their lab.

In 7.4% of papers ( $n = 4$ ), authors described employing the *Wizard and Oz model*, a human-centric method that uses real technology in a real environment. For example, Iwamura et al. (2011) conducted a field experiment in a supermarket with a semi-autonomous robot that provides shopping assistance to older adults. They suggest their robot can be fully autonomous in the near future.

In 7.4% of papers ( $n = 4$ ), authors described employing the *Oz with Wizard model*, a robot-centric method where users are involved in a WoZ experiment but not precisely measured. For example, Noda et al. (2007) describe an experiment testing a haptic interaction classification method at the Osaka Science Museum, allowing participants to freely interact with their robot. The participants themselves are not measured, but how robot classifies their actions is.

In 1.9% of papers ( $n = 1$ ), authors described employing the *Oz of Wizard model*, a robot-centric

method focused on “advancing enabling technologies.” In this model, the user is simulated (or minimally involved), and their performance is tightly intertwined ultimately with system performance.

No studies employed the *Wizard Nor Oz model*, a method in which all aspects of the system are simulated. This is explained by the fact that our inclusion criteria explicitly specified a physically embodied robot interacting with a real human.

## 4. Discussion

### 4.1 How WoZ is currently used in HRI research

*4.1.1 Types of wizard control.* For the papers included in this review, HRI researchers used WoZ primarily for natural language processing and non-verbal behavior simulation (72.2% and 48.1% respectively). Wizards also often performed navigation and mobility tasks (29.6%); while manipulation, sensing, and mapping simulations were far less commonly employed (3.9%, 1.9%, and 1.9% respectively).

This usage pattern suggests that the truly “hard problems” in HRI still lie in verbal and visual human language understanding, requiring the use of a Wizard. Lee (2003) suggested that natural language processing is “AI-complete”, meaning that it is one of the most challenging problems in artificial intelligence. This sentiment has recently been echoed by Picard (2009a, 2009b), a pioneer in affective computing, who stated that machine understanding of facial expressions alone is one of the hardest, most complex problems in computer science. Thus, we will likely see Wizards being used for these tasks for the foreseeable future.

*4.1.2 Fraser and Gilbert (1991) criteria.* Generally, authors clearly articulated that their WoZ simulation was feasible to simulate (92.6%), convincing (72.2%), and specifiable (94.4%).

However, given the majority of papers used WoZ for fairly complex verbal and visual human language understanding, it is unclear if these systems were truly specifiable. While the majority of papers constrained what the Wizard could produce (90.7%), very few constrained what the Wizard could recognize (11%). Further, of those papers that did constrain Wizard production, very few researchers had Wizards deliberately make machinelike errors simulating a future system (3.7%), and very few actually measured Wizard error (3.7%).

However, even if an HRI researcher creates recognition constraints and deliberate production errors for a Wizard, a savvy Wizard (and User) could likely overcome them. For example, even with simulating an occluded visual sensor, a degraded audio signal, and system lag, a Wizard could still ‘thin slice’ the situation to know what action to take (Ambady & Rosenthal, 1992), whereas this would be an incredibly difficult, if not impossible feat for a robot (Riek & Robinson, 2011). Nonetheless, controlled Wizard recognition and production errors could still help to address some of the aforementioned WoZ methodological concerns.

Another way researchers can ensure a specifiable system is to provide sufficient training to their Wizards. Only 3 papers (5.4%) actually described Wizard training; it is unclear if the experiments from the other 51 papers actually provided training, or if it was in any way controlled or considered by the researchers. It is also not clear what the Wizards knew in advance of the experiment in terms of its hypotheses, robot behavioral expectations, etc. These too may be influencing the experimental results.

*4.1.3 Green et al. (2004) criteria.* Overall, the included papers followed criteria by Green et al. (2004) closely, by specifying what users could do during the interaction and providing them with a clear scenario (94.4%), specifying their behavioral hypotheses for the user (79.6%) and for the robot (98.1%).

It would be helpful if authors were more explicit about their behavioral hypotheses or predictions of the user, as often this was not easy to determine, particularly in the robot-centric papers. This lack of clear reporting makes future replication more difficult.

**4.1.4 Kelley (1984) criteria.** Only a quarter of papers (24.1%) clearly described their WoZ simulation as being part of an iterative design process. It is unclear from the remaining papers if researchers were indeed performing iterative WoZ simulations and just not describing doing so, or if they were declaring their reported results as final. Also, authors did not always mention their use of WoZ as a limitation when reporting their results.

One thing that would be helpful is for authors to clearly sketch out their iterative plans for using WoZ control for their robot. Doing so would help other HRI researchers either replicate or carry on the work. Also, it would be useful for authors to mention the limitations of their results, given they are a part of an iterative methodology.

**4.1.5 Steinfeld et al. (2009) criteria.** Authors predominantly used the *Wizard of Oz* model (68.5%), a human-centric method using possibly very futuristic technological simulation on the robot. This is unsurprising given the majority of papers using WoZ control for human language understanding.

Only 7.4% of papers employed the *Wizard and Oz* model, a human-centric method that uses real technology in a real environment. This is unfortunate given Steinfeld et al. suggest this is the ‘preferred method’ of all six types of WoZ simulations (Steinfeld et al., 2009). However, given the inherent challenges in conducting real-world HRI studies with real robots (Burke et al., 2011), it is not surprising.

## 4.2 Limitations and Future Work

This systematic review has a few limitations. First, due to the fact that the review spanned a decade of literature, a few tradeoffs had to be made regarding inclusion criteria. By only including ACM and IEEE peer-reviewed publications, we excluded papers from other venues that publishes HRI work in our community, such as publications from professional societies such as the American Association for the Advancement of Artificial Intelligence (AAAI) and the Human Factors and Ergonomic Society (HFES), as well as journals such as the Journal on Social Robotics and the Interaction Studies Journal. A broader search may have yielded different findings, and is a reasonable next step for future work in this area.

Secondly, by only including publications written in English, we may have excluded relevant publications from other countries’ professional societies and journals that publish HRI work in other languages. Hopefully HRI researchers who speak these other languages fluently can replicate this systematic review and include those publication venues to see if the same trends appear there.

Third, with regard to the Steinfeld et al. criteria, the use of the term “Wizard of Oz” in our initial search may have inadvertently biased our selection against “Oz of Wizard” papers. Papers of this sort (primarily robot-centric) are less likely to use the term “Wizard of Oz.” In the future, a broader search which includes more keywords robot-centric papers are likely to use (e.g., “remote operator”) would help address this.

Finally, these findings will be most applicable to those who wish to contribute generalizable knowledge to the HRI community through the use of WoZ as an experimental technique. For those researchers who only intend to use WoZ as an early prototyping design technique, where unscripted, uncontrolled wizard operation is acceptable, these findings may be less applicable.

Table 2: Proposed reporting guidelines for HRI studies that use WoZ.

Experimental Component	Question
Robot	How many robots were used?
	What kind(s) of robot(s)? (e.g., humanoid, zoomorphic, mechanical, android?)
	What level(s) of autonomy? (i.e., Which components of the robot(s) were autonomous and which were controlled by the wizard?)
	What were the robot's capabilities?
	What hypotheses did the researcher have for the robot?
User	How many users participated in total, and per experimental trial?
	What were the user demographics, sampling procedure, etc.?
	What instructions were provided to the user?
	What behavioral hypothesis does the researcher have about the user?
	Was the simulation convincing to the user?
	What expectations did the user have about the robot, before and after the experiment?
Wizard	How many wizards were used?
	What were the wizard demographics? (e.g., the researcher, lab mates, naïve?)
	Did the wizard know the behavioral hypothesis of the experiment?
	What were the wizard production variables and how were they controlled for?
	What were the wizard recognition variables and how were they controlled for?
	How did the experimenter control for wizard error (deliberate and accidental)?
	How much and what sort of training did wizards receive prior to starting the experiment?
General	Where did the experiment take place?
	What were the environmental constants and how were they controlled?
	What scenarios did the researchers employ?
	Was this experiment part of an iterative design process?
	Does the paper discuss the limitations of WoZ?

### 4.3 New WoZ reporting guidelines for HRI

In the field of biomedical research, reporting guidelines “specify a minimum set of items required for a clear and transparent account of what was done and what was found in a research study, reflecting, in particular, issues that might introduce bias into the research” (U.S. National Library of Medicine, 2011).

Thus, in order to encourage easy replication and peer-assessment of WoZ studies, we suggest the following reporting guidelines, shown in Table 2. These guidelines are based on our results as well as the aforementioned WoZ and HRI literature. Though it is not critical that every single HRI WoZ paper details every single one of these criteria, it will be helpful if researchers at least carefully consider these questions.

The table is divided up into four experimental components (Robot, User, Wizard, and General), each with affiliated questions to motivate thinking about the topic. The Robot component asks questions about which robots were used, their capabilities, and the researcher’s predictions for what the robot will do during the experiment. These questions were primarily motivated by Green et al. (2004).

The User component asks questions about the users interacting with the robot, and includes questions regarding their demographics, the instructions they were provided, experimental hypotheses about their behavior, the user’s expectations about the robot, and the user’s perceptions of the WoZ simulation. These questions were motivated by Green et al. (2004), Dahlbäck et al. (1993), Fraser and Gilbert (1991), and Lohse (2010).

Next, the Wizard component asks questions about the Wizard operating the robot, and includes questions concerning the wizard’s demographics, prior knowledge of the experimental hypotheses, identification and control of wizard production / recognition variables, how the researcher controlled wizard errors, and quantity and type of training the wizard received. These questions were motivated by Fraser and Gilbert (1991) and Steinfeld et al. (2009).

Finally, the General section includes questions regarding the experimental environment, experimental scenarios, whether the experiment was part of an iterative design process, and, finally, if the paper discussed the limitations of using WoZ. These questions were motivated by Kelley (1984), Fraser and Gilbert (1991), Green et al. (2004), Dahlbäck et al. (1993), and Steinfeld et al. (2009).

### 4.4 Summary

We presented a systematic review of the HRI literature between 2001 - 2011 that included the use of the WoZ experimental technique. Using methodological criteria proposed by Fraser and Gilbert (1991), Green et al. (2004), Steinfeld et al. (2009), and Kelley (1984), 54 papers were analyzed on how they controlled for various aspects of system specification and experimental design, as well as how clearly they presented their results to the scientific community. Our key findings are summarized below:

*Types of Wizard control.* The majority of experiments used a Wizard to control for natural language (72.2%) and non-verbal behavioral (48.1%) processing. Wizards also often performed navigation and mobility tasks (29.6%); while manipulation, sensing, and mapping simulations were far less commonly employed.

*Wizard constraints and training.* Most papers reported constraining what Wizards could produce on the robot (90.7%), but very few reported constraining what the Wizard could recognize (11%). Even fewer papers reported measuring Wizard error (3.7%), and few described giving the Wizards training before the experiment (5.4%).

*User scenario constraints.* Most papers clearly specified giving users a scenario, as well as specifying both their user and robot behavioral hypotheses.

*WoZ paradigms employed.* Authors predominantly used the *Wizard of Oz* model (68.5%), a human-centric method using possibly very futuristic technological simulation on the robot; only 7.4% used the preferred *Wizard and Oz* model, a human-centric method using real technology in a real environment.

*Iterative design.* Only a quarter of included papers clearly mentioned their use of WoZ was part of an iterative design process.

Based on these findings, we presented new reporting guidelines with the aim of guiding researchers toward more rigorous means of conducting and reporting HRI WoZ experiments, ideally helping circumvent the existing methodological concerns surrounding this methodology.

## Acknowledgements

We would like to acknowledge the assistance of Michael Gonzales in figure preparation, and the reviewers for their helpful comments.

## References

*An asterisk before an entry indicates it was included in the systematic review.*

ACM. (2012). *ACM Digital Library*.

\*Alissandrakis, A., & Miyake, Y. (2009). Human to robot demonstrations of routine home tasks: Adaptation to the robot's preferred style of demonstration. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 141–146). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326253>

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2), 256–274. <http://dx.doi.org/10.1037/0033-2909.111.2.256>

\*Austermann, A., & Yamada, S. (2007). “Good robot, bad robot” – Analyzing users’ feedback in a human-robot teaching task. In *17th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 41–46). IEEE. <http://dx.doi.org/10.1109/ROMAN.2008.4600641>

\*Bainbridge, W. A. A., Hart, J., Kim, E. S., & Scassellati, B. (2008). The effect of presence on human-robot interaction. In *17th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 701–706). IEEE. <http://dx.doi.org/10.1109/ROMAN.2008.4600749>

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. <http://dx.doi.org/10.1109/IROS.2005.1545011>

\*Burghart, C., Holzapfel, H., Haeussling, R., & Breuer, S. (2007). Coding interaction patterns between human and receptionist robot. In *7th IEEE-RAS International Conference on Humanoid Robots*. IEEE. <http://dx.doi.org/10.1109/ICHR.2007.4813909>

Burke, J. L., Christensen, H. I., Menassa, R., Koeppe, R., Dyer, J., & Munich, M. (2011). HRI: The Real World. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (p. 103). New York, New York, USA: ACM. <http://dx.doi.org/10.1145/1957656.1957681>

\*Chao, C., Cakmak, M., & Thomaz, A. L. (2010). Transparent active learning for robots. In *5th ACM/IEEE International Conference on Human-Robot Interaction*. <http://dx.doi.org/10.1145/1734454.1734562>

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of internal medicine*, 126(5), 376–380.

Dählback, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, 6(4), 258–266. [http://dx.doi.org/10.1016/0950-7051\(93\)90017-N](http://dx.doi.org/10.1016/0950-7051(93)90017-N)

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., & Gandy, M. (2005). Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing*, 4(4), 18–26. <http://dx.doi.org/10.1109/MPRV.2005.93>

- EPSRC. (2011). *Principles of robotics: Regulating Robots in the Real World* (Tech. Rep.). Engineering and Physical Science Research Council, EPSRC.
- \*Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)*. ACM. <http://dx.doi.org/10.1145/1957656.1957672>
- \*Fischer, K., & Lohse, M. (2007). Shaping naive users' models of robots' situation awareness. In *16th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 534–539). IEEE. <http://dx.doi.org/10.1109/ROMAN.2007.4415144>
- Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, 5(1), 81–99. [http://dx.doi.org/10.1016/0885-2308\(91\)90019-M](http://dx.doi.org/10.1016/0885-2308(91)90019-M)
- \*Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., Kitamura, Y., & Tsujino, H. (2008). Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI)*. ACM. <http://dx.doi.org/10.1145/1452392.1452452>
- \*Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *12th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)*. IEEE. <http://dx.doi.org/10.1109/ROMAN.2003.1251796>
- Goodrich, M. A., & Schultz, A. C. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1(3), 203–275. <http://dx.doi.org/10.1561/11000000005>
- \*Green, A. (2009). Towards a minimal interaction scheme: Initial observations on gesture repertoire and communication strategies. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)*. IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326202>
- \*Green, A., Eklundh, K. S. S., & Wrede, B. (2006). Integrating miscommunication analysis in natural language interface design for a service robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. <http://dx.doi.org/10.1109/IROS.2006.282256>
- Green, A., Hüttenrauch, H., & Eklundh, K. S. (2004). Applying the Wizard-of-Oz framework to cooperative service discovery and configuration. In *13th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 575–580). IEEE. <http://dx.doi.org/10.1109/ROMAN.2004.1374824>
- \*Harriott, C. E. E., Zhang, T., & Adams, J. A. A. (2011). Evaluating the applicability of current models of workload to peer-based human-robot teams. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (pp. 45–52). ACM. <http://dx.doi.org/10.1145/1957656.1957670>
- \*Hato, Y., Kanold, T., Ishii, K., & Imai, M. (2009). Showing awareness of humans' context to involve humans in interaction. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 663–668). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326283>
- \*Hato, Y., Satake, S., Kanda, T., Imai, M., & Hagita, N. (2010). Pointing to space: Modeling of deictic interaction referring to regions. In *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 301–308). IEEE. <http://dx.doi.org/10.1109/HRI.2010.5453180>
- \*Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2006). The influence of a robot's social abilities on acceptance by elderly users. In *15th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 521–526). IEEE. <http://dx.doi.org/10.1109/ROMAN.2006.314442>
- \*Heerink, M., Krose, B., Evers, V., & Wielinga, B. (2007). Observing conversational expressiveness of elderly users interacting with a robot and screen agent. In *IEEE 10th International Conference on Rehabilitation Robotics (ICORR)* (pp. 751–756). IEEE. <http://dx.doi.org/10.1109/ICORR.2007.4428509>
- \*Hüttenrauch, H., Eklundh, K. S. S., Green, A., & Topp, E. A. A. (2006). Investigating spatial relationships in human-robot interaction. In *Proceedings of the IEEE/RSJ International Conference on intelligent robots and Systems (IROS)*. IEEE. <http://dx.doi.org/10.1109/IROS.2006.282535>
- IEEE. (2012). *IEEEExplore*.
- \*Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H., & Hagita, N. (2011). Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (pp. 449–456). ACM. <http://dx.doi.org/10.1145/1957656.1957816>

- \*Jones, T., Lawson, S., & Mills, D. (2008). Interaction with a zoomorphic robot that exhibits canid mechanisms of behaviour. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 2128–2133). IEEE. <http://dx.doi.org/10.1109/ROBOT.2008.4543521>
- Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems*, 2(1), 26–41. <http://dx.doi.org/10.1145/357417.357420>
- \*Kim, E., Leyzberg, D., Tsui, K. M., & Scassellati, B. (2009). How people talk when teaching a robot. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 23). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1514095.1514102>
- \*Kim, H., Kwak, S., & Kim, M. (2008). Personality design of sociable robots by control of gesture design factors. In *17th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 494–499). IEEE. <http://dx.doi.org/10.1109/ROMAN.2008.4600715>
- \*Kim, T., & Hinds, P. (2006). Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction. In *15th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 80–85). IEEE. <http://dx.doi.org/10.1109/ROMAN.2006.314398>
- \*Koay, K. L., Dautenhahn, K., Woods, S. N., & Walters, M. L. (2006). Empirical results from using a comfort level device in human-robot interaction studies. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI)* (p. 194). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1121241.1121276>
- \*Komatsu, T., & Kuki, N. (2009). Investigating the contributing factors to make users react toward an on-screen agent as if they are reacting toward a robotic agent. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 651–656). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326350>
- \*Kuzuoka, H., Suzuki, Y., Yamashita, J., & Yamazaki, K. (2010). Reconfiguring spatial formation arrangement by robot body orientation. In *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 285–292). IEEE. <http://dx.doi.org/10.1109/HRI.2010.5453182>
- \*Lang, C., Hanheide, M., Lohse, M., Wersing, H., & Sagerer, G. (2009). Feedback interpretation based on facial expressions in human-robot interaction. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 189–194). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326199>
- Lee, L. (2003). "I'm sorry Dave, I'm afraid I can't do that": Linguistics, Statistics, and Natural Language Processing circa 2001. *Computer Science: Reflections on the Field, (report of the National Academies' Study on the Fundamentals of Computer Science)*, 7.
- \*Lee, M. K. K., Forlizzi, J., Rybski, P. E. E., Crabbe, F., Chung, W., Finkle, J., et al. (2009). The Snackbot: Documenting the design of a robot for long-term human-robot interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 7–14). ACM. <http://dx.doi.org/10.1145/1514095.1514100>
- Lohse, M. (2010). *Investigating the influence of situations and expectations on user behavior: empirical analyses in human-robot interaction*. Unpublished doctoral dissertation, Doctoral Thesis. Bielefeld University, Technical Faculty, Germany.
- \*Looije, R., Cnossen, F., & Neerinx, M. A. A. (2006). Incorporating guidelines for health assistance into a socially intelligent robot. In *15th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 515–520). IEEE. <http://dx.doi.org/10.1109/ROMAN.2006.314441>
- Miller, K. W. (2010). It's Not Nice to Fool Humans. *IT Professional*, 12(1), 51–52. <http://dx.doi.org/10.1109/MITP.2010.32>
- \*Mohammad, Y., Nishida, T., & Okada, S. (2009). Unsupervised simultaneous learning of gestures, actions and their associations for human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2537–2544). IEEE. <http://dx.doi.org/10.1109/IROS.2009.5353987>
- \*Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., & Hagita, N. (2011). Effect of robot's active touch on people's motivation. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (pp. 465–472). ACM. <http://dx.doi.org/10.1145/1957656.1957819>
- \*Noda, T., Miyashita, T., Ishiguro, H., Hagita, N., Engineering, I., Laboratories, C., et al. (2007). Map



- acquisition and classification of haptic interaction using cross correlation between distributed tactile sensors on the whole body surface. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1099–1105). IEEE. <http://dx.doi.org/10.1109/IROS.2007.4399573>
- \*Okita, S. Y. Y., & Sarvadevabhatla, R. (2009). Learning together: ASIMO developing an interactive learning partnership with children. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 1125–1130). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326135>
- \*Okuno, Y., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2009). Providing route directions: design of robot's utterance, gesture, and timing. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 53–60). ACM. <http://dx.doi.org/10.1145/1514095.1514108>
- \*Osawa, H., Ohmura, R., & Imai, M. (2008, December). Embodiment of an Agent by Anthropomorphization of a Common Object. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (Vol. 2, pp. 484–490). IEEE. <http://dx.doi.org/10.1109/WIIAT.2008.129>
- \*Otero, N., Alissandrakis, A., Dautenhahn, K., Nehaniv, C., Syrdal, D. S. S., & Koay, K. L. L. (2008). Human to Robot Demonstrations of Routine Home Tasks: Exploring the Role of the Robots Feedback. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 177–184). ACM. <http://dx.doi.org/10.1145/1349822.1349846>
- \*Perzanowski, D., Brock, D., Adams, W., Bugajska, M., Schultz, A. C., Trafton, J. G., et al. (2003). Finding the FOO: A pilot study for a multimodal interface. In *IEEE International Conference on Systems, Man and Cybernetics* (Vol. 4, pp. 3218–3223). IEEE. <http://dx.doi.org/10.1109/ICSMC.2003.1244386>
- Picard, R. W. (2009a). *Emotion Technology: From Autism to Customer Experience and Decision Making*. Microsoft Research Cambridge.
- Picard, R. W. (2009b, December). Future affective technology for autism and emotion communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535), 3575–3584. <http://dx.doi.org/10.1098/rstb.2009.0143>
- \*Ralph, M., & Moussa, M. (2008). Toward a natural language interface for transferring grasping skills to robots. *IEEE Transactions on Robotics*, 24(2), 468–475. <http://dx.doi.org/10.1109/TRO.2008.915445>
- Riek, L. D., & Robinson, P. (2011). Challenges and Opportunities in Building Socially Intelligent Machines. *IEEE Signal Processing*, 28(3), 146–149. <http://dx.doi.org/10.1109/MSP.2011.940412>
- Riek, L. D., & Watson, R. (2010). The Age of Avatar Realism. *IEEE Robotics & Automation Magazine*, 17(4), 37–42. <http://dx.doi.org/10.1109/MRA.2010.938841>
- \*Robins, B., Dautenhahn, K., Nehaniv, C. L. L., Mirza, N., Francois, D., & Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. In *IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 716–722). IEEE. <http://dx.doi.org/10.1109/ROMAN.2005.1513864>
- \*Robins, B., Dautenhahn, K., Te Boekhorst, R., & Nehaniv, C. L. (2008). Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinesics. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 17–24). ACM. <http://dx.doi.org/10.1145/1349822.1349826>
- \*Rosenthal, S., Dey, A. K. K., & Veloso, M. (2009). How robots' questions affect the accuracy of the human responses. In *18th IEEE International Symposium on Robots and Human Interactive Communication (Ro-MAN)* (pp. 1137–1142). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326291>
- \*Rouanet, P., Danieau, F., & Oudeyer, P. Y. Y. (2011). A robotic game to evaluate interfaces used to show and teach visual objects to a robot in real world condition. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI)* (pp. 313–320). ACM. <http://dx.doi.org/10.1145/1957656.1957782>
- \*Ryokai, K., Lee, M. J., & Breitbart, J. M. (2009). Children's storytelling and programming with robotic characters. In *Proceedings of the 7th ACM Conference on Creativity and Cognition (C&C)* (p. 19). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1640233.1640240>
- Scholtz, J. (2003). Theory and evaluation of human robot interactions. In *Proceedings of the 36th annual Hawaii International Conference on System Sciences*. IEEE. <http://dx.doi.org/10.1109/HICSS.2003.1174284>
- \*Shadid, S., Krahmer, E., Swerts, M., & Mubin, O. (2010). Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience. In *Proceedings of the 22nd Conference of*

- the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI)* (pp. 332–335). <http://dx.doi.org/10.1145/1952222.1952294>
- \*Shimada, M., Yoshikawa, Y., & Ishiguro, H. (2009). Social balancing effect of eye contact. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 1119–1124). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326147>
- \*Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H., & Hagita, N. (2007). Group attention control for communication robots with wizard of OZ approach. In *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 121). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1228716.1228733>
- \*Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C. T. T., Ishiguro, H., & Hagita, N. (2008). A Semi-autonomous communication robot - A field trial at a train station. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 303–310). ACM. <http://dx.doi.org/10.1145/1349822.1349862>
- \*Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., & Hagita, N. (2008). How quickly should communication robots respond? In *Proceedings of the 3rd International Conference on Human-Robot Interaction (HRI)* (p. 153). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1349822.1349843>
- \*Siino, R. M., Chung, J., & Hinds, P. J. (2008). Colleague vs. tool: Effects of disclosure in human-robot collaboration. In *17th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 651–656). IEEE. <http://dx.doi.org/10.1109/ROMAN.2008.4600725>
- Steinfeld, A., Jenkins, O. C., & Scassellati, B. (2009). The Oz of Wizard: Simulating the human for interaction research. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*. ACM. <http://dx.doi.org/10.1145/1514095.1514115>
- \*Tanaka, K., Zuo, X., Sagano, Y., & Oka, N. (2007). Learning the meaning of action commands based on "no news is good news" criterion. In *Proceedings of the 2007 Workshop on Multimodal Interfaces in Semantic Interaction (WMISI)* (pp. 9–16). New York, New York, USA: ACM Press. <http://dx.doi.org/10.1145/1330572.1330574>
- U.S. National Library of Medicine. (2011). *Research Reporting Guidelines and Initiatives: By Organization*. U.S. National Library of Medicine.
- \*Walters, M. L., Dautenhahn, K., Kaouri, C., Boekhorst, R., Nehaniv, C., Werry, I., et al. (2005). Close encounters : Spatial distances between people and a robot of mechanistic appearance. In *5th IEEE-RAS International Conference on Humanoid Robots* (pp. 450–455). IEEE. <http://dx.doi.org/10.1109/ICHR.2005.1573608>
- Weiss, A. (2010). *Validation of an Evaluation Framework for Human-Robot Interaction. The Impact of Usability, Social Acceptance, User Experience, and Societal Impact on Collaboration with Humanoid Robots*. Unpublished doctoral dissertation, University of Salzburg.
- \*Weiss, A., Igelsbock, J., Calinon, S., Billard, A., & Tscheligi, M. (2009). Teaching a humanoid: A user study on learning by demonstration with HOAP-3. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 147–152). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326274>
- \*Wendt, C. S., & Berg, G. (2009). Nonverbal humor as a new dimension of HRI. In *18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 183–188). IEEE. <http://dx.doi.org/10.1109/ROMAN.2009.5326230>
- \*Xu, Y., Guillemot, M., & Nishida, T. (2007). An experimental study of gesture-based human-robot interface. In *IEEE/ICME International Conference on Complex Medical Engineering (CME)* (pp. 457–463). IEEE. <http://dx.doi.org/10.1109/TRO.2009.2035747>
- \*Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2010). A Model of Proximity Control for Information-Presenting Robots. *IEEE Transactions on Robotics*, 26(1), 187–195. <http://dx.doi.org/10.1109/TRO.2009.2035747>

---

Laurel D. Riek, Dept. of Computer Science and Engineering, University of Notre Dame, Notre Dame, USA. Email: lriek@nd.edu