Considerations for Minimizing Data Collection Biases for Eliciting Natural Behavior in Human-Robot Interaction

Snehesh Shrestha
snehesh@umd.edu
University of Maryland College Park
College Park, Marland, USA

Ge Gao
gegao@umd.edu
University of Maryland College Park
College Park, Marland, USA

Cornelia Fermuller
fermulcm@umd.edu
University of Maryland College Park
College Park, Marland, USA

Yiannis Aloimonos
jyaloimo@umd.edu
University of Maryland College Park
College Park, Marland, USA

Figure 1: A sample of participants with diverse hand gestures while they command the robot to cut onions.

ABSTRACT

Many of us researchers take extra measures to control for known-unknowns. However, unknown-unknowns can, at best, be negligible, but otherwise, they could produce unreliable data that might have dire consequences in real-life downstream applications. Human-Robot Interaction standards informed by empirical data could save us time and effort and provide us with the path toward the robots of the future. To this end, we share some of our pilot studies, lessons learned, and how they affected the outcome of our experiments. While these aspects might not be publishable in themselves, we hope our work might save time and effort for other researchers towards their research and serve as additional considerations for discussion at the workshop.

KEYWORDS

hri, human factor, robots, data collection, standards, pilot studies

1 EXTENDED ABSTRACT

Natural human-robot interaction data can help robots learn from signals that are unstructured, mixed-modal, and consists of implied contexts. In our experiments, we find that they contain contradictory phrases and repair mechanisms. To incite natural human behavior in the lab is challenging. We took inspiration from prior literature, conducted several pilot studies, and developed a Wizard of Oz study (WoZ) experiment design to incite natural emergent human behavior.

To achieve the best of both worlds (in the wild and controlled lab study), our WoZ experiment design deceived the participants into believing the remote-controlled robot was fully autonomous. In this extended abstract, we will discuss a number of selected factors that we considered. These are some pilot studies we conducted that could affect participant behavior to validate independent and dependent control variables as well as the workflow. We will discuss the following considerations whose findings informed our experiment design decisions.

- The effect of experiment *instructions*
- The *WoZ clues* that participants might be able to use to figure out the hidden agenda
- The *priming* effects from *practice* sessions
- The *background noise*, and
- The robot's *appearance* and *identity*

1.1 Instructions

We tested various modalities for our applications based on the recommendations [4]. Our findings in our pilot studies were in line with [2, 4], where the instruction modality significantly impacted the participants' behavior. For instances when text instructions were provided similar to [1], participants preferred speech and used the exact words for the action and the object with little or no gestures. With videos of people performing the task similar to [2], participants copied the exact style of the demonstration of the actor. The one with the most variance in speech vocabulary and styles of gestures was when we showed before-after video clips to show the pre-task and post-task states; for example, to turn on a stove, we showed a zoomed-in video of a stove that was turned off and faded out to a video of the stove with the fire burning. For cutting an apple, a video of a whole apple on a cutting board being approached by a knife and faded into the apple that was cut into pieces where the knife is leaving the screen. And these videos were repeated in a loop with a 1-second gap. With such substantial differences in behavioral outcomes, we believe the instruction mechanisms used in the experiments and data collection sessions could benefit from guidelines and standardization.

1.2 WoZ Clues

Wizard of Oz study (WoZ) is research experiment method where participants interact with a system the participants believe to be autonomous, however, the system is being operation fully or partially by another human[3, 6]. However, in WoZ studies, people can intuitively figure out the patterns, such as key press and mouse click sounds corresponding to robot actions. We experimented with masking the actual clicks and key presses with random ones. However, in the post-interview, the pilot test participants still seem to be able to figure out that researchers might be controlling the robot. So we created soft rubber remote control keys that use an IR receiver using an Arduino micro-controller USB adapter to send keys to the WoZ UI with virtually no sound that the researcher kept in their pocket. With this implementation, during the experiment, the researchers made sure when the experiment was being conducted, they did not sit at the control computer and appear to be moving around doing other things, appearing busy, staring at their phone, seemingly distracted, or looking at the participants showing attention in making sure the system was working without any technical issues. With this implementation, all of our participants believed that the robot was acting independently, and none suspected the WoZ setup to be a possibility.

1.3 Practice Session

Practice sessions, especially performed right before the experiments, can strongly impact the outcome of the participant's behavior. For example, in our experiment, it was essential to ensure that participants were not primed to use one modality versus the other. So steps were taken to design the session with a mixture of related and unrelated commands where both speech and gestures were used to command the robots. If participants used a single modality only, they were encouraged to test out using the other modality. Participants interacted with the robot and asked researchers questions during practice. Once the practice was completed, participants were not allowed to interact with anyone other than the robot, even if they had questions or felt stuck as they were told that the experiment was designed for them to experience such scenarios and had to use creative methods to make the robot understand what they wanted the robot to do.

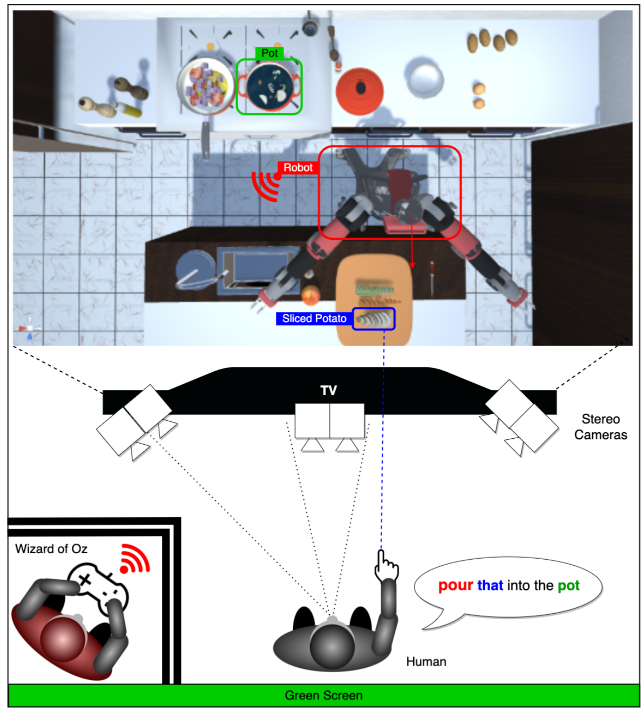

Figure 2: The researcher (WoZ) shown here in the bottom-left box is controlling the robot hidden from the participant, who believes the robot is autonomous.

1.4 Background Noise

One hypothesis was that background noise can cause people to use more gestures. We considered three types of noise recording playback (lawn mower, people talking, and music) but only tested with people talking as background noise as that was the only example people found to be believable and not simulated. We tested three sets of loudness (M (dB) = 58, 63, 70, SD = 10, 13, 15). In our study (N=8), from people's use of speech and gesture and the post-interview, we found that (a) people tune out the background noise instead of using more gestures, (b) people wait for gaps of silence or lower-level noise in cases of speech or periodic noise, and (c) the noise had to be so loud that none of the speech could be heard for them to use gestures instead of speech. For these reasons, we decided not to use background noise as an independent variable. However, as a future possible directions, with a visualization of robot's perception of the sound, for example, robot picking up words from the background noise, or participant's speech being drowned out by the background noise, participants might use more gestures.

1.5 Robot Appearance and Identify

To reduce the effect of perceived gender, age, and personality by manipulating facial attributes, we considered the 17 face dimensions based on [5] study to design the face of the robot to be the most neutral face. The mouth of the robot was removed as not having a mouth did not have a significant adverse effect on the neutral perception of the robot. Having a mouth gave people the idea that the robot could speak, potentially causing the participant to prefer speech over gesture. To appear dynamic, friendly, and intelligent, we made the robot blink randomly between 12 and 18 blinks per minute [7] with ease-in and ease-out motion profile [8, 9]. We further conducted pilot tests to analyze the head nod motions (velocity and the number of nods) and facial expressions for confusion expression. Additionally, we avoided using gender-specific pronouns "he/him" and "she/her" and referred to the robot as "the robot" or "Baxter," which is also the manufacturer-given name printed on the robot body that tends to be used both as a male and female name [10].

REFERENCES

[1] Jessica R Cauchard, Jane L E, Kevin Y Zhai, and James A Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 361–365.

[2] Emiko Charbonneau, Andrew Miller, and Joseph J LaViola. 2011. Teach me to dance: exploring player experience and performance in full body dance games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (Lisbon, Portugal) *(ACE '11, Article 43)*. Association for Computing Machinery, New York, NY, USA, 1–8.

[3] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*. 193–200.

[4] Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing people for training gestural interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1737–1746.

[5] Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the Design Space of Rendered Robot Faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 96–104.

[6] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.

[7] Kazuki Takashima, Yasuko Omori, Yoshiharu Yoshimoto, Yuichi Itoh, Yoshifumi Kitamura, and Fumio Kishino. 2008. Effects of avatar's blinking animation on person impressions. In *Graphics Interface*. researchgate.net, 169–176.

[8] Frank Thomas, Ollie Johnston, and Frank Thomas. 1995. *The illusion of life: Disney animation*. Hyperion New York.

[9] Laura C Trutoiu, Elizabeth J Carter, Iain Matthews, and Jessica K Hodgins. 2011. Modeling and animating eye blinks. *ACM Trans. Appl. Percept.* 8, 3 (Aug. 2011), 1–17.

[10] Wikipedia contributors. 2021. Baxter (name). https://en.wikipedia.org/wiki/Baxter_(name). Accessed: 08-01-2021.