“Hi human, can we talk?” An in-the-wild study template for robots approaching unsuspecting participants

Elaheh Sanoubari, James E. Young

Abstract— **We present a study design template for conducting an in-the-wild human-robot interaction study, where a robot has to “cold-call” approach and attempt to start a conversation with unsuspecting participants, ultimately asking the participant for help with a task. Our protocol uses the length of time that a person helps the robot as a measure of engagement; a researcher can modify their interaction protocol and investigate how this impacts the engagement time, in addition to any qualitative investigation employed. We further provide an ethics protocol solution that addresses concerns over informed consent, while balancing for minimal impact on natural interaction by relying primarily on post-test debriefing. We envision that this template can be used by researchers as a starting-point for developing similar in-the-wild studies, providing base-line ethics and interaction solutions from which additional variables and conditions can be added.**

I. Introduction

The current generation of robots is increasingly moving into human-centric everyday life scenarios, including museums [1,7–9], shopping malls [2,4,6], and train stations [5,14]. Key to this area of research is in-the-wild studies, where robots are studied directly within ecologically valid, real-world scenarios where they interact with the people already in those spaces [12]. In comparison to more traditional laboratory studies, which provide control and consistency but result in artificial interactions, in-the-wild research inverts this balance: interactions are more natural and valid, but the lack of control of the research environment introduces challenges. We provide a template study design that addresses typical challenges that we experienced conducting an in-the-wild study, and can be used as a starting point for developing similar studies.

Our study design specifically targets “cold calling” robots: those that approach and engage unsuspecting people (who did not agree to be participants). Further, our robot attempts to get help from a participant, and encourages ongoing engagement with mild persuasion. Thus, this paper provides templates for initiating interaction, and for encouraging ongoing engagement, where the duration of interaction (how long the person helps the robot) varies and reflects a measure of engagement. We envision that other researchers could build on this template by adding their own variables and manipulations.

One key problem with conducting in-the-wild studies is receiving institutional research ethics approval: requirements of informed consent, particularly for protocols that involve deception, can impose cumbersome constraints (such as pre-interaction interviews and consent forms) that hamper natural interaction and thus lesson the desired research validity with in-the-wild studies. We present our solution (approved by our local board) that includes an informed consent procedure that primarily involves post-interaction researcher engagement, minimizing the impact on the validity of the experiment.

We developed this study for our project on culture-based communication styles: we investigate the impact of a robot’s communication style on its ability to get help from strangers. Specifically, we evaluate the impact of culturally-aligned versus misaligned language on how much a person is willing to help a robot in passing, measured by how long they help the robot. Importantly, participants simply believe that they are

E. S. Author, and J. Y. Author are with the Department of Computer Science, University of Manitoba, Winnipeg, MB, R3T 2N2 Canada (e-mail: elaheh.sanoubari@cs.umanitoba.ca , young@cs.umanitoba.ca).

Figure 1. A participant interacting with our robot, Sam, from the camera’s view (left) and the robot’s view (right). Note the brochures in a basket attached to the robot.

helping a robot with a task, requiring deception. We used a Double 2 telepresence robot controlled using the Wizard of Oz technique, using voice synthesis: we led participants to believe the robot to be autonomous (Figure 1). We gave the robot an animated face, and named it the unisex “Sam”. Our study took place in public indoor spaces around our university campus.

This research posed specific challenges in designing the study protocol, challenges which we believe are common to many in-the-wild studies. In this paper, we provide some details on these challenges, and provide template solutions that other researchers can build from to design in-the-wild studies.

II. Research Ethics Boards: Informed Consent and Debriefing Challenges for In-the-Wild Studies

We faced challenges obtaining approval from our institution’s research ethics board for our in-the-wild study. The common requirement of informed consent requires us to clearly inform participants of the study details, and to receive their explicit consent, in order to use their data – typically, such consent is received prior to participation. Initially, our ethics board requested us to obtain this prior participant consent, for all who directly interact with the robot, as well as passersby who may be caught in a video feed. Unfortunately, this level of pre-study interaction with participants would negatively impact research validity from an in-the-wild perspective.

It is common to study people in public settings without receiving informed consent, using the argument that people have a reasonable expectation of being observed. However, our ethics board highlighted that our introduction of a manipulation (the robot and its behavior) could not be reasonably expected by people, and thus this argument was not acceptable.

To complicate informed consent, our study involves deception (common in social HRI): participants believe that the robot is autonomous and requires help with a task, but actually we use the Wizard of Oz technique, and study the impact of speaking style on interaction. Modern ethics standards require us to fully debrief participants on deception as soon as possible after interaction. However, immediate debriefing in-the-wild may inform other passersby (potential participants), whether done verbally by the robot, or in writing (e.g., on screen or with pamphlets): a participant may react, verbally noting study details (e.g., robot is not intelligent). Our research ethics board initially requested this typical immediate debriefing, and raised valid concern over any delay in conveying this information.

A further component of informed consent standards is to enable participants to retroactively withdraw consent, particularly important when deception is involved. Once a participant is debriefed, they may regret their participation, and we must enable them to withdraw consent and have data destroyed.

Our solution to these challenges, cooperatively designed and accepted by our research ethics board, has four components:

- provide clear public awareness that a study is taking place, while detracting attention from the true study purpose,
- clearly advertise multiple means for people to contact researchers to ask questions or have their data destroyed,
- delay informed consent for participants until after interaction, destroying data when consent is not received,
- create a space near (but separated from) the experiment where participants get debriefed and provide consent.

While we concede that this approach is no longer perfectly in-the-wild, as participants are aware that a study is taking place, they are not aware of the study purpose. We see this as a reasonable compromise given the research goals (ecologically valid interaction) and constraints of modern ethics protocols. Below we provide the details of this approach.

a) Provide Public Awareness While Detracting Attention from True Study Purpose

We placed posters prominently in the study area (Figure 2) to inform passersby that a study was taking place, and put information brochures on a basket on the robot itself. With this material, we were careful not to include any study details (e.g., the project title), and focused instead on the key ethics-relevant points of communication: that a study was taking place, and that the space was being video recorded for research purposes.

Given our goal to minimize participants guessing that they (or their interactions) were being tested, we designed our media to suggest that the robot was being tested. On all materials we titled the research "Sam: the robot", prominently featuring Sam's animated face (Figure 3). Further, when a participant engaged the robot (see next section), it gave a back-story of how it was trying to improve its algorithms. This placed the focus on the robot and its capabilities, and not the interaction.

Overall, this strategy informed passersby of the study without providing detail on the actual study purpose.

b) Clear Mechanisms for Contacting Researchers

We provided a range of contact information (email, phone, fax) on all media for the researchers and research ethics board secretariat, if people had any concerns or questions, or wanted their data (video, any interaction, etc.) destroyed. This mitigated issues with not having pre-study informed consent, by providing clear means for anyone to opt out. Further, with using posters and brochures we did not require a researcher to be in the interaction space, increasing in-the-wild validity.

We did have a researcher on-site (10-20m away, passively observing, Figure 2), wearing a large ID badge (with robot face Figure 3) for easy identification. Further, the robot could direct people to the researcher if they had concerns or questions.

c) Post-Interaction Informed Consent

Consent procedures for in-the-wild studies, including study description, debriefing, and getting signatures, should happen after interaction to protect study validity. While it is common in studies involving deception to first receive prior consent, and provide a later means to withdraw it (after debriefing), we feel that any prior interaction with researchers would greatly reduce the validity of the in-the-wild methodology.

One challenge we found was compelling participants – who just finished interacting with a robot – to take more time to engage a researcher, and read and fill forms and questionnaires. In many cases, participants would simply walk away from the robot – in this case, we do not have consent and must destroy the data. Our solution was, as soon as it was clear the person was leaving, for the robot to show a picture of a gift card with text asking them to engage the researcher. The researcher (10-20m away, Figure 2) was easily identifiable (Figure 3).

An additional challenge in the wild is ensuring that consent is associated with the correct participant. Participants come and go with little structure, for example, they may approach the researcher out of turn, and while debriefing one participant, another may interact with the robot. People may also ask for a gift card, without having interacted with the robot. Our solution was to add a *completion code*, a series of seemingly 6 random numbers and letters (e.g., DJNAJ6) to the gift card screen. Participants needed to note (or take a picture) of this code to receive their gift card. We facilitated this by providing small cards and pencils in the robot's basket, alongside the brochures.

We note an alternative method considered: an online system that a participant could engage using their completion code for access and data association. However, there was concern over the delay before debriefing happened, as well as identification: we had no way to verify if the person using the completion code was the person who interacted with the robot.

Our post-interaction debriefing method, away from the study space, and using a completion code, results in a stringent enforcement of informed consent with minimal impact on the research environment, at the cost of more strictly discarding data for participants who do not complete the whole procedure.

d) Near-Immediate Debriefing in a Semi-Private Space

By conducting debriefing and informed-consent procedures (consent form, etc.) at a separate nearby space (Figure 2), we reduced the potential for debriefing to inform other potential participants on the study, either through the debriefing itself or actions of the participant being debriefed. The delay introduced

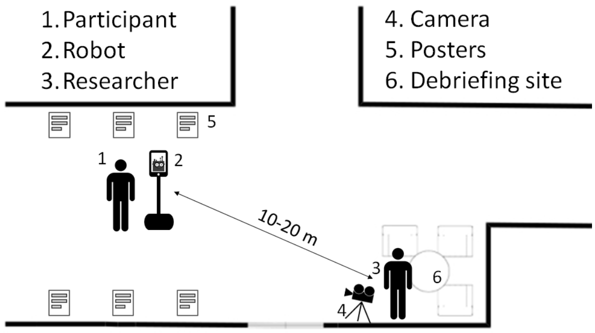

Figure 2. Diagram of the experiment space

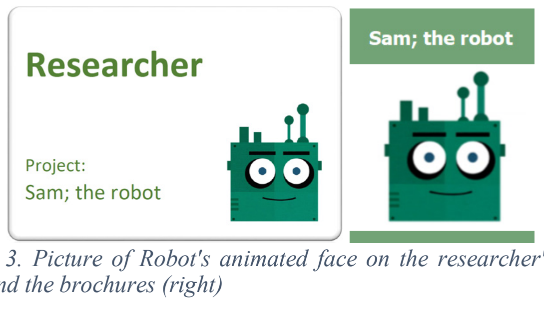

Figure 3. Picture of Robot's animated face on the researcher's ID (left) and the brochures (right)

between the end of interaction, and debriefing, was minimal, and we compensated for this extra required step by providing a gift-card incentive. Given that we further remove data for those who not engage this process, this is considered a reasonable compromise from an ethics perspective.

This space was also semi-private, with chairs for seating and a small table, which provided a comfortable space for the debriefing and study-related discussion.

Having the researcher separated from the robot (10-20m, Figure 2), not directly watching interaction, helped create a natural setting of being unobserved. The researcher is still close enough to approach a participant in case of problem, or for a participant to find easily if they have any questions.

Our method provides awareness of the study but through passive mediums (posters, brochures), provides various means for anyone to contact the researchers, has a researcher on-site but separated from the robot, and delays informed consent until after interaction, but is strict on discarding data without consent. Overall, this provides a balance between the modern research standards of informed consent, and the ecological validity demands of in-the-wild research studies.

III. A Study Template for In-the-Wild Cold Calling Robot Studies

We provide a template for in-the-wild cold-calling robots that other researchers can use as a starting point for similar studies. We first introduce the design from a high level, and follow with specific challenges (i.e., initiating and maintaining interaction) faced. We provide the full study script at the end of the paper.

Within this study design, researchers can introduce their own manipulations (e.g., social cues, changes in robot morphology, change in communication protocol, etc.) and measures. The study design itself includes a variable-length interaction component, which can serve as one measure of engagement.

The flow of our template is provided in Figure 4. The robot attempts to start unsolicited interaction by approaching passersby and asking for help. If the person declines, the interaction ends. If they agree to help, the robot starts by asking questions – an analog to a demographics questionnaire.

The robot then provides a short backstory. Researchers can use this to set a stage or purpose, and is designed to increase engagement (grab interest), increase believability [15], and initiate a more involved social interaction with the participant. In our version, the robot states that they want to improve their algorithms and wants help from the person in order to do that.

Following, in the next phase the robot asks for help from the person with a task, and attempts to maintain interaction: it only ends when the participant explicitly stops helping. In our case, the robot asked for help with classic image labelling questions, where the person answered a yes-or-no question about whether a given image contains a given object (Figure 5). We created a database of images from crowdsource.google.com, manually filtered for general-audience appropriateness.

It is important to note that this is a distractor task – we do not suggest for researchers to investigate people's ability with image labelling. Instead, this provides an excuse for people to interact with the robot, while a researcher can introduce other manipulations and measures, around this task. We selected this task as a believable problem for machines, but generally reasonably easy for people to accomplish. At the same time, it is not inherently enjoyable and we do not expect participants to continue the task for long periods of time.

The robot keeps asking until the participant indicates they are done, verbally, or by starting to walk away. At this point the robot requests the person to continue. If that fails, we execute the end-study protocol with the gift card, completion code, and debriefing. The duration of interaction varied between participants, and can be used as a measure of engagement.

a) Initiating Interaction with Potential Participants

In piloting, we faced challenges with initiating interaction with participants, who were busily passing by. Unlike some studies where a robot has a clearly-defined utility role (e.g., a museum guide [1]), a cold-calling robot has the challenge of engaging busy people (a problem previously noted [6]).

Informally, we noticed that some people appeared to be intimidated, unsure what was happening, and others ignored the robot (a problem noted by others [11,13]). We designed our posters and brochures to clearly indicate that a new robot was being tested, to help people understand the context.

Our specific protocol, which we found to have reasonable success (that is, we can get participants), was for the robot to first physically move toward people, and home in on anyone who looks at or pays attention to the robot (done via Wizard of Oz). The robot then tries to initiate conversation with a simple greeting, and if the person responds, follows up with a request

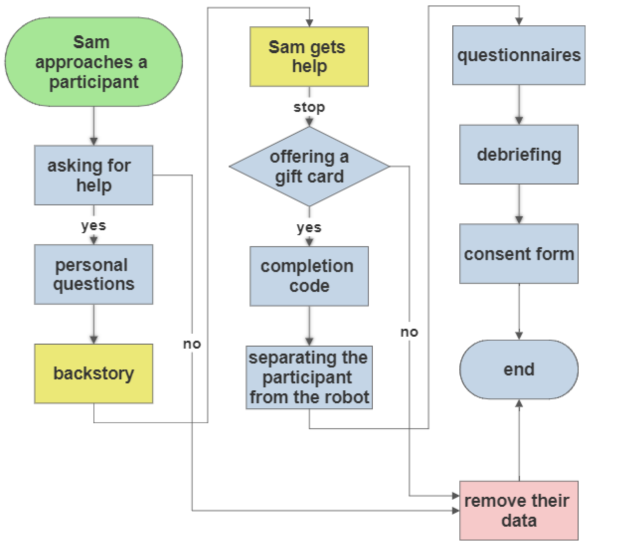

Figure 4. Flow diagram of the study procedure, starting with the robot approaching a participant. Note that in several cases participant research data is removed.

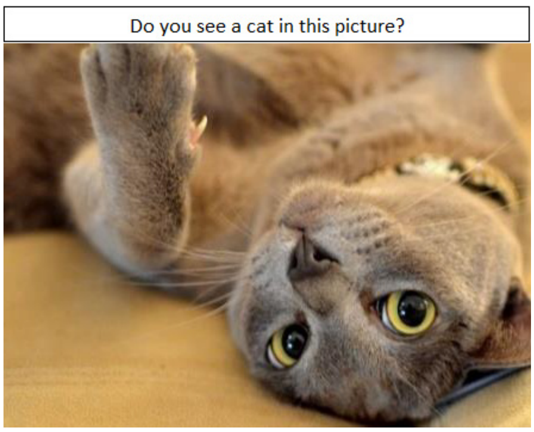

Figure 5. Example of the image labeling question used in the experiment. https://crowdsource.google.com/

for help. In our experience, the physical act of moving towards potential participants was helpful for triggering interest; informally, we note that being approached by someone may create social pressure to interact with them. In our experiment, one participant told the researcher (during debriefing) that he had been watching the robot for a while but was not sure if he was allowed to approach it, because it looked expensive.

b) Maintaining Interaction

We faced challenges maintaining interaction once started. Our general strategy was to build a believable character (shown to help in engagement [10]) that aimed to build a positive rapport in order to exert social pressure (e.g., as in [16]). For example, Sam's face was designed to be welcoming and kind, Sam asked the person's name to build intimacy, and humbly asked for help. Throughout interaction Sam would make empathic statements such as "Do you know how that feels?" (regarding not wanting to bother lab mates).

Building from our own social HRI experiment experience, to maintain believability we carefully avoided the robot appearing too intelligent or clever. During pilots, when an operator ad-libbed a portion of the interaction, some would question the scenario, for example, stating "there is someone behind this." Instead, we pre-scripted the entire conversation tree. When participants changed the subject, asked probing questions (such as the robot's favorite colour), or tried to be humorous, the robot would simply state that it did not understand, and returned to the stock interaction. This technique has been shown to be also successful with other work, to annoy participants [3], consistent with our experience.

Inversely, during pilot studies we noted that operators who were less proficient and operating the robot, and thus were slow to respond and poor at driving toward people, appeared to reduce believability compared with experienced operators.

c) Evaluation

Unfortunately, at this point we do not have formal evaluation of either our ethics procedure or our study template design. As with many social HRI experiment designs, we approached this pragmatically and analytically, refining through trial and error, and pilot studies. Our work has both successfully passed our local ethics board standards, and, has been used in an actual study (forthcoming). Despite the lack of detailed evaluation, we hope that our explanations, both for ethics and study design, in addition to our particular solutions, will be valuable to other researchers working on similar studies.

IV. Conclusion

In this paper we presented a template study design for in-the-wild cold-calling robot studies, including a detailed study flow and script, with an interaction-initiation protocol and distractor task. We also provide a discussion on challenges with maintaining research ethics standards and proposed a set of solutions for in-the-wild work. This study design has been piloted and successfully used in an actual study, for initiating and maintaining interaction with passersby, where an interaction manipulation resulting in a change in interaction duration.

We believe that this discussion and study design can serve as a starting point for other researchers to develop their own in-the-wild studies. It can easily be adapted and modified, with new manipulations and measurements introduced.

Study Script

Here we include the study script as used and tested in our actual study. Study phases are indicated.

[Sam approaches robot]
Hi there! Can I talk to you?
I am Sam. What is your name?

[Asking for help]
Would you help me?
I am here to improve my algorithms and communication skills.

[Personal Questions]
Can I ask a few questions about you, first?
Are you currently a student?
Which country's culture do you identify closest to?
What is your favorite color?

[Backstory]
Well, I am one of the first robots that is allowed to go around the university!
I don't like to constantly ask for help from others in the lab. You know how that feels?
Anyway, I want to be independent from others, so I decided to improve my algorithms by volunteering for this experiment.

[If agreed to help]
Perfect! Your answers will help me stand on my own feet.

[Distractor Task]
Please answer the following questions by saying Yes, No, or Skip.

[The following responses chosen in order after every image labelling instance]
Thanks! Here is the next one.
Thank you! Your answers help me be more independent.
Sweet! And this one?
With your answers, I will be able to get better results on my own.
Next question?
I am impressed!
What about this one?
You're on fire dude! What about this one?
You're the best! Next?
Your answers help me stand on my own feet!
You're on a roll!
Now I can be on my own in the lab!
Interesting!

[First persuasion]
Can you please continue? I need more data.

[Second persuasion]
But how do I become independent without your help? Please help me more!

[Study end]
Alright. Thank you for your help.
Please find the researcher at [location], who will thank you with a 10$ gift card if you fill out a questionnaire about your experience.
Make sure you have your completion code! Have a good day!

References

1. Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke. 2005. Towards a humanoid museum guide robot that interacts with

multiple persons. *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots* 2005: 418–423. https://doi.org/10.1109/ICHR.2005.1573603

2. Chaitanya P. Gharpure and Vladimir A. Kulyukin. 2008. Robot-assisted shopping for the blind: Issues in spatial cognition and product selection. *Intelligent Service Robotics* 1, 3: 237–251. https://doi.org/10.1007/s11370-008-0020-9

3. Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz, and Jue Wang. 2005. Designing robots for long-term social interaction. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*: 2199–2204. https://doi.org/10.1109/IROS.2005.1545303

4. H. M. Gross, H. Boehme, Ch Schroeter, S. Mueller, A. Koenig, E. Einhorn, Ch Martin, M. Merten, and A. Bley. 2009. TOOMAS: Interactive shopping guide robots in everyday use - Final implementation and experiences from long-term field trials. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*: 2005–2012. https://doi.org/10.1109/IROS.2009.5354497

5. K Hayashi and Daisuke Sakamoto. 2007. Humanoid robots as a passive-social medium-a field experiment at a train station. *Human-Robot Interaction*: 137–144. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6251680

6. Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2009. An affective guide robot in a shopping mall. *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09*: 173. https://doi.org/10.1145/1514095.1514127

7. Y Kuno, K Sadazuka, M Kawashima, K Yamazaki, A Yamazaki, and H Kuzuoka. 2007. Museum guide robot based on sociological interaction analysis. *Proceedings of the SIGCHI conference on Human factors in computing systems*, September 2015: 1194. https://doi.org/10.1145/1240624.1240804

8. I.R. Nourbakhsh, C. Kunz, and T. Willeke. 2003. The mobot museum robot installations: a five year experiment. *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)* 4, October: 3636–3641. https://doi.org/10.1109/IROS.2003.1249720

9. Illah R. Nourbakhsh, Judith Bobenage, Sebastien Grange, Ron Lutz, Roland Meyer, and Alvaro Soto. 1999. An affective mobile robot educator with a full-time job. *Artificial Intelligence* 114, 1–2: 95–124. https://doi.org/10.1016/S0004-3702(99)00027-2

10. Raquel Ros, Marco Nalin, Rachel Wood, and Paul Baxter. 2011. Child-robot interaction in the wild: advice to the aspiring experimenter. ICMI '11 Proceedings of the 13th international conference on multimodal interfaces 335–342. https://doi.org/10.1145/2070481.2070545

11. Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, Christian Becker-Asano, Kohei Ogawa, Shuichi Nishio, and Hiroshi Ishiguro. 2014. The Uncanny in the Wild. Analysis of Unscripted Human-Android Interaction in the Field. *International Journal of Social Robotics* 6, 1: 67–83. https://doi.org/10.1007/s12369-013-0198-7

12. Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. *International Workshop on Advanced Motion Control, AMC* 2006: 576–581. https://doi.org/10.1109/AMC.2006.1631758

13. Alessandra Maria Sabelli, Takayuki Kanda, and Norihiro Hagita. 2011. A conversational robot in an elderly care center: An ethnographic study. *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*: 37–44. https://doi.org/10.1145/1957656.1957669

14. Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2011. Field trial of a networked robot at a train station. *International Journal of Social Robotics* 3, 1: 27–40. https://doi.org/10.1007/s12369-010-0077-4

15. Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee, Imran Fanaswala, Brett Browning, Jodi Forlizzi, and Majd Sakr. 2011. Believable Robot Characters. *AI Magazine* 32, 4: 39. https://doi.org/10.1609/aimag.v32i4.2383

16. Sebastian Thrun, Jamie Schulte, and Chuck Rosenberg. 2000. Interaction With Mobile Robots in Public Places. *IEEE Intelligent Systems*: 7–11. Retrieved from http://www.cs.cmu.edu/~thrun/movies/papers/thrun.robointeract00.pdf