Toward a Science of Robotics: Goals and Standards for Experimental Research

Leila Takayama
Willow Garage
68 Willow Road, Menlo Park, CA 94025
Email: takayama@willowgarage.com

I. Introduction

The fusion of science and technology is a challenge faced by many engineering research efforts, including robotics. One particularly useful frame for thinking about the relationships between science and technology was well articulated by Michael Polanyi, a philosopher who coined the terms "empirical technology" and "systematic technology" [3]. Empirical technologies are those technologies that are largely unscientific, but are so unique and innovative that they inspire new scientific theories and explorations, thereby contributing to science. Systematic technologies are those that are deeply informed by current scientific knowledge, thereby benefiting from science. In an ideal world, any given robotic system would contain the best of both–incorporating the state of the art in scientific knowledge and inspiring fundamentally new research questions and directions of inquiry. This balance of science and technology facilitates a reciprocally generative feedback cycle between the engineering work of inventing new technologies, and the scientific work of generating new knowledge.

One of the challenges of doing rigorous scientific research in robotics is creating a body of knowledge through proper scientific experimentation. This paper aims to take a concrete step toward articulating the goals and standards necessary for meeting this challenge.

II. Goals and Standards

One goal may be to build robotic systems that are both systematic and empirical technologies in order to develop a science of robotics rather than just explore robotics for the sake of explorin g robotics. In other words, these robotic systems should be deeply informed by existing scientific knowledge, and also inspire new directions of scientific inquiry. To this end, it is informative to review the standards by which scientific experiments are designed, conducted, and presented.

Among the many standards of scientific research is the expectation that phenomena are observable, repeatable, and objective (or, at least, free of as many sources of bias as possible). The implications of these standards are that experiments must (1) use good measures, which are standardized and/or usable by others, and are as objective as possible, (2) clearly and thoroughly explain the exact methods used to run the study so that others can replicate the study to see if they get similar or different results, and (3) minimize the subjectivity in the experimental observations, e.g., using double-blind experimental set ups in which neither the experimenter nor the participant (if there is one) knows what experimental condition is being run. Many of the lessons learned from methods courses in experimental social science research are directly applicable to the methods that could improve the quality of experimental research in the robotics community.

III. Variable-Based Research

Scientific experiments are typically informed by theories, hypotheses, or just a hunch that there is some variable (X) that causally influences another variable (Y). These experimental results allow for generating knowledge about how to change X in order to influence Y in a desirable direction. An extremely powerful and applicable explanation of how this notion applies to the study of technologies is Nass's argument for taking a variable-based approach to technologically-oriented research [2]. Nass argues that comparing technology A (e.g., radio) to technology B (e.g., television) does not make sense unless you only care about the difference between the specific technologies; this is A/B testing. Instead, in order to make more generalizable conclusions, it is even more powerful to figure out exactly which *dimensions* of a technology matter (e.g., audibility, size) in terms of influence on the outcomes that one cares about (e.g., performance on task). This variable-based approach is one that has been applied to human-computer interaction and human-robot interaction (HRI) to create knowledge that is usable and testable by others in the future, regardless of changes in the specific technologies. To ground these ideas in real experimental robotics work, the following section presents two recently published studies that take a variable-based approach to HRI.

IV. Examples of Variable-Based Research

These two studies present concrete examples of published HRI research. Both studies were designed to be informative for HRI theories and the design of human-robot interactions, thereby benefiting both scientific inquiry and engineering / design work in HRI. They employed random assignment of participants to experiment conditions, isolating the causal independent variables of the study designs. We used between-participant experiment designs, meaning that any given participant was unaware of what variables were being manipulated in the study at the time of participation. Only statistically significant differences between experiment conditions were reported as results. Because we used standard data analysis techniques (e.g., analysis of variance), the results lend themselves to analytical scrutiny and comparison against related work. None of the participants were labmates or close friends of the experimenters of the study so they were not particularly inclined to help the experimenters find particular results. A limitation of these studies is that they used a university student populations so the results might only generalize to other university student populations. However, this is explicitly mentioned in the text of the reports and can be addressed by replicating the studies with other populations. All limitations we could identify were included in the text of these publications so that others can decide how to interpret the findings and how to improve their own related research in future work.

A. Study 1: Self Extension Into Robots

Taking a variable-based approach to HRI, the first study [1] investigated the research question: How does robot form and experience with building a robot influence a person's experience with the robot and perceptions of the robot? To address this question, we designed a 2 (robot form: humanoid vs. car) x 2 (assembler: self vs. other) between-participants experiment (*N*=56) in which people built a robot (either a humanoid or car) and used a robot (either the one they built or one that was supposedly built by someone else) to play a game. We evaluated people's perceptions of how much the robot's personality overlapped with their own personality as a proxy for understanding how much people experienced a sense of self-extension into the robot. We found that people had more positive experiences and felt a greater sense of self-extension into the robots that they built themselves. Similarly, car-like robots evoked more positive experiences and a greater self-extension into the robots when compared to human-like robots.

Two experiment design decisions are worth noting here. As much as possible, we held constant all variables that were not of interest. This is consistent with the notion of ceteris paribus in experimental psychology, i.e., manipulating only the variables one is studying and holding all else constant. For example, we made sure that the humanoid and car-like robot forms were similar in size, and were made of the same materials. As with all of our lab's studies, we used standardized questionnaires or slight modifications thereof so that our results would be usable by others and more readily compared against existing literatures.

B. Study 2: Disagreeing Robots

Taking a variable-based approach to HRI, the second study [4] investigated the research questions: How does robot disagreement with a person influence performance upon a human-robot collaborative task? How does the placement of robot voices (speakers) influence the human-robot team performance? To address these questions, we designed a 2 (robot disagreement: none vs. some) x 2 (robot voice location: on robot body vs. in control box) between-participants experiment (*N*=40) in the context of a human-robot desert survival task. We measured decision-making outcomes and attitudes toward the robot in this study. We found that people are accepting of robot's disagreements with them and that they are more accepting of disagreeing robots if their voices are placed on separate bodies (e.g., control boxes).

A couple of experiment design decisions were important here. First, we chose a slightly modified version of a standard experimental task (desert survival collaborative decision-making task) that is actually used in real-world training settings and has been used across dozens of studies in human-computer interaction in the past. This makes the findings more comparable to previous work. Second, we used both behavioral and attitudinal measures in our study so that conclusions could be drawn about both performance on the task, and feelings about the experience.

V. Conclusion

Good experimental research requires having the humility to give oneself the chance to be wrong. Testing theories and iterating upon them is the nature of scientific research. Empirically generated research findings are beneficial to scientific knowledge in that they either support one's ideas when the results are consistent with one's predictions, or they guide those ideas in new directions when the results are not consistent with one's predictions. By clearly and thoroughly describing one's experimental research, other scientists can run identical or similar studies to test the validity and generalizability of one's claims.

While generalizability in robotics is difficult to achieve due to the diversity of robotic parts and systems in various research labs, it is not impossible. If one lab does studies on their unique robot, the hypotheses, methods, and measures used by that lab are still usable by others. By testing and replicating others' evaluations of their own robots, the field can move forward to create more general results that apply to multiple types of robots instead of finding particular results that uniquely apply to particular robots. In an effort to improve the sharing of such research across labs, Willow Garage is generating and sharing an open source robot operating system (ROS) and will soon be launching a beta program to make a hardware robotic platform available to approximately ten robotics labs so that they can share and test their work more readily.

References

[1] V. Groom, L. Takayama, and C. Nass, "I Am My Robot: The Impact of Robot-Building and Robot Form on Operators," *Proceedings of the Human-Robot Interaction Conference: HRI 2009*, pp. 31-36.

[2] C. Nass, L. Mason, "On the Study of Tehcnology and Task: A Variable-Based Approach," in *Organization and Communication Technology*, Ed. J. Fulk and C. Steinfeld, Newbury Park: Sage, 1990, pp. 46-67.

[3] M. Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy*, Chicago: University of Chicago Press, 1990.

[4] L. Takayama, V. Groom, and C. Nass, "I'm Sorry, Dave: I'm Afraid I Won't Do That: Social Aspects of Human-Agent Conflict'" *Proceedings of Human Factors in Computing Systems: CHI 2009*, April 2009, pp. 2099-2107.