The Science of Human-Robot Interaction

SARA KIESLER, Carnegie Mellon University
MICHAEL A. GOODRICH, Brigham Young University

Researchers in the field of human-robot interaction (HRI) are celebrating the transition to an official ACM Transactions journal. It is a significant achievement for HRI. This field has faced many challenges, including scarce resources for conducting HRI research, the difficulty of conducting the research itself—integrating hardware and software; building robots that can interact with people; piecing together the intricacies of seeing, thinking, and acting when a person encounters a robot; understanding and shaping ethical norms and standards; and stepping behind the mythologies that have long governed cultural attitudes for and against robots. HRI are presented here.

Even as HRI becomes more mature, the core disciplines of the field face many controversies and subtleties that complicate our research questions and activities. Here's an example: It was long thought that human beings can discriminate only the seven so-called universal emotions of surprise, happiness, sadness, anger, disgust, fear, and contempt in others' faces and, therefore, that modeling this ability in a robot would be soon possible. But now, psychologists have reported that people can make much finer-grained emotional distinctions, raising the bar for designing a robot. Advances in technology, machine learning, and data science also present higher standards for what a robot can do with, and for, people.

What Do We Need to Learn About Humans? At its core, the science of HRI is a science of interaction, but we cannot assume that the state of being human is a known entity that does not change. We still need to learn more about the following:

- the causes and manifestations of people's intentions, so that robots can ascertain them
- trust, and what it takes in different cultures to develop and destroy trust

---

Authors' addresses: S. Kiesler, Human Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213; email: kiesler@cs.cmu.edu; M. A. Goodrich, Computer Science Department, Brigham Young University, 3361 TMCB, Provo, UT 84602; email: mike@cs.byu.edu.

- how to provide people with insight into artificial intelligence, so that mutual understanding is possible
- how to develop norms of human-robot collaboration that correspond in usable ways with norms of human-human collaboration
- how different groups, cultures, and individuals at different ages and with different capabilities and resources, can benefit from, or be harmed from, human-robot interactions.

What Do We Need to Learn About Algorithms that Work with Humans? In some ways, this question is the opposing partner of the first. If the science of HRI is a science of interaction, then the development of embodied algorithms should seek not to advance autonomy *per se* but rather to enable sustainable and productive interaction. What representations are most compatible with how a human might reason about a problem (e.g., cognitive robotics)? What algorithms and interface modes are needed to enable humans to understand the glut of data that will be returned from ubiquitous and sensor-rich robots? How can models of self-interested rationality and precise reasoning be modified to enable a human to accurately infer robot intent, to make robot behaviors transparent and predictable, and to avoid inducing relationship-marring miscalibrations of trust?

How Can We Scale Interactions? Key questions moving forward are what can, will, and should happen when many robots and people interact over many months or many years. Longitudinal interactions call for new models and methods of interaction, artificial intelligence that predicts future expectations and behavior, human and machine learning, and a deeper understanding of human relationships with technology. We also need to learn more about how robots with different capabilities, functions, and representations can work with each other as well as with people.

What are Successful Human-Robot Interactions? Many researchers are interested in the area of work called "HRI for social welfare" and hope that HRI can contribute to a better world. Every issue in *JHRI* has papers reflecting that goal. Yet there remain steep obstacles. Algorithms are often not accountable to the people that they affect. Privacy and security lag behind, especially in the domain of the Internet of Things. Robots are, for the most part, not sustainable. And HRI seems not to be much attuned to our multicultural, diverse society. Thus, plenty of obstacles exist to make social welfare HRI a difficult goal.

Can We Apply New Scientific Methods? The development of computational social science, smart and learning algorithms, big data, and the availability of robots for experimentation now makes it possible to study HRI-related phenomena that, up to now, have forced researchers to use small samples and unreliable robots. Now, we can study factory automation, consumer robot appliances, robotics in hospital services, disaster robots, shopping robots, and toy robots with larger and more diverse samples in diverse domains, along with varied data sources, to provide more generalizable knowledge than was possible before. Machine-learning and other advanced methods can reveal patterns in the brain, in social networks, and across social media that were otherwise hidden, giving us a better ability to plan and implement smart and connected robots and robotic devices. HRI will have far more impact as a field if researchers tackle more varied and longer-term outcomes—outcomes over months or years, not hours or days—and measure factors such as cultural background or social connectivity that intervene or moderate these outcomes.

Evolving computational techniques, modeling methods, and models of human understanding and behavior require new methods in HRI. New methods are required for validating start-to-end systems for designing and verifying algorithms that are resilient to relationship and technology failures. Formal proofs will be needed, but proofs of small-scale properties, such as the completeness of a path-planning algorithm, will be less important than proofs about interactional equilibria under various perturbation conditions. User studies will be needed, but

studies with small sample sizes in laboratory conditions with college-aged participants will need to be augmented with longitudinal studies and supported by ethnographic methods. New [1, 3, 4] establish the importance, validity, and relevance of technology design and human studies will need to be developed to enable longitudinal HRI to move forward. New methods will be needed for performing the data science to store, parse, interpret, and use the data that will be gathered in longitudinal interactions. We will also need to study the secondary impact of [2, 5] change: for instance, how robots at work affect people's social relationships and lives.

REFERENCES

[1] Samuel Arbesman. 2012. *The Half-life of Facts: Why Everything We Know Has an Expiration Date*. Current, New York, NY.

[2] R. C. Granger and R. Maynard. 2015. Unlocking the potential of the "what works" approach to policy making and practice: Improving impact evaluations. *American Journal of Evaluation* 36, 558–569.

[3] Donald O. Hebb. 1975. *Einführung in Die Moderne Psychologie*. Beltz, Weinheim, Germany.

[4] Larry Laudan. 1981. A confutation of convergent realism. *Philosophy of Science* 48, 1, 19–49.

[5] D. T. Miller, J. E. Dannals, and J. J. Zlatev. 2017. Behavioral processes in long-lag intervention studies. *Perspectives on Psychological Science* 12, 3 (2017), 454–467.