EDITORIAL

Introduction to the Special Issue on Psychological Benchmarks of Human–Robot Interaction

Karl F. MacDorman and Peter H. Kahn, Jr.
School of Informatics, Indiana University / Department of Psychology, University of Washington

The idea for this special issue took shape during discussions on the prospects for using technology to simulate nature and, in particular, the human form. Could it be possible to devise an artificial human being? The computer scientist and robotic engineer, with such ambitions, can reply: "Sure, just give us major funding, say a half billion dollars, and 30 years, and we'll show you how." The skeptic can reply, "You're kidding, right?" Along these lines, debates have raged in the philosophy of mind and cognitive science on whether anything like present day computers could implement a *conscious* mind, one that could experience *what it's like* to be human. Moreover, because it is unclear even what makes *us* conscious, this problem is likely to remain a hard one for years to come.

Nevertheless, we can reframe what a human being is from the standpoint of human attribution. If the technologist insists that it will eventually be possible to build an artificial human being, it is important to determine what would count as one in our own estimation, taking a view from the outside. The question then becomes: What are the benchmarks — categories of interaction that capture fundamental aspects of human life — by which we could measure progress toward this goal? Getting the right set of benchmarks then becomes critical for the emerging field of human–robot interaction (HRI). The benchmarks can help establish the questions the field asks in setting its research agenda, determining where funding is directed, and shaping how graduate students are educated. The right set of benchmarks will also be important to other disciplines, such as comparative psychology, and to meeting the long-term needs of society in areas such as nursing, eldercare, and social work.

To these ends, Peter H. Kahn, Jr. and his colleagues proposed six benchmarks in a paper he showed Kerstin Dautenhahn, who was then organizing the 15th IEEE International Symposium on Robot and Human Interactive Communication

Interaction Studies 8:3 (2007), **359–362**.
ISSN 1572–0373 / E-ISSN 1572–0381

(September 6 to September 8, 2006, University of Hertfordshire, Hatfield, United Kingdom). Dautenhahn suggested organizing special sessions at the symposium to explore this issue. A number of new benchmarks were proposed by participants who held sometimes divergent sets of assumptions. For example, drawing on work in phenomenology, Christopher H. Ramey proposed *conscience* as a benchmark for designing social robots. Michelle B. Cowley proposed *intent* as exhibited during strategic interactions as a benchmark. Karl F. MacDorman and Stephen J. Cowley proposed the ability to maintain *long-term relationships* as a benchmark for robot personhood. At the symposium these researchers not only proposed new benchmarks but also critiqued each other's.[1] That spirit continues in this special issue.

We begin with the updated proposal for nine psychological benchmarks by **Peter H. Kahn, Jr., Hiroshi Ishiguro, Batya Friedman, Takayuki Kanda, Nathan G. Freier, Rachel L. Severson, and Jessica Miller.** Their benchmarks comprise *autonomy, imitation, intrinsic moral value, moral accountability, privacy, reciprocity, conventionality, creativity,* and *authenticity of relation.* This paper has the qualities of a target article given that it has served as a point of reference for the articles that follow. Next **Justine Cassell and Andrea Tartaro** propose *intersubjectivity* as a benchmark for human–agent interaction: Do people respond to the agent, consciously and subconsciously, as if it were human, attributing similar intentions? Cassell and Tartaro take issue with the current emphasis on believability in HRI and, in some quarters, the emphasis on reproducing human physical appearance. What is more important to them is whether the microstructure of human responses to the agent match those directed toward other human beings. In this same vein, **Billy Lee** devises a set of nonverbal behavior codes for concrete, observable actions that occur between couples and between strangers. These codes provide act-specifications for a human–robot interaction benchmark for *intimacy*. Lee proposes that a tele-operated android designed with sensitivity to these codes might afford a feeling of presence lacking in telephone counseling, thus enhancing therapeutic outcomes.

David Feil-Seifer, Kristine Skinner, and Maja J. Matarić consider appropriate benchmarks for socially assistive robots working, for example, in education, rehabilitation, nursing, and care for the elderly. The 10 benchmarks they propose — *safety, scalability, autonomy, imitation, privacy, understanding of domain, social success,* and *impact on the user's care, life and caregivers* — are divided among the general areas of robot technology, social interaction, and social success. In addition to measuring a robot's psychological impact on individual users, they frame their benchmarks in terms of the achievement of the robot's intended task and how the robot influences the social dynamics of its user community. **Sylvain Cali-**

non and Aude G. Billard propose benchmarks to improve methods of programming robots by teaching, including the relative importance of different *sensory modalities* and the usefulness of robot *skill rehearsal* in noticing the robot's current understanding of the task.

Our next two papers likewise examine the role of human mental models in collaborating with robots. **Debra Bernstein, Kevin Crowley, and Illah Nourbakhsh** propose *relationship potential* as an HRI benchmark: the potential for robot and human to develop a long-term collaboration. Drawing on a study of children's development of mental models when interacting with a robot rover, they point out that success largely hinges on people being able to develop accurate beliefs about how robots work. For many kinds of collaboration, it is counterproductive to obscure the robot's internal workings by prompting users to view the robot through a misleading human metaphor. Instead, the robot interface should clearly communicate its capabilities and limitations. In turn, **Victoria Groom and Clifford Nass** express skepticism about the prospect for building *robot teammates* in a controversial paper. They consider robots to lack two essential prerequisites for teamwork: a sense of self and humanlike mental models. Thus, they think it unlikely robots will be able to satisfy the benchmarks they propose for teammates. They propose guidelines for other organizational structures that could enable people to work alongside robots. Given that the field of human–robot interaction is quite new and current attempts at building human–robot teams have used robots that are at best semi-autonomous, we consider the "the jury to be still out" on the prospects of building robot teammates.

In our final paper, **Sherry Turkle** offers a critique of psychological benchmarks based on decades of ethnographic work in human–computer and human–robot interaction. Her studies reveal people's capacity *to nurture* their digital companions, to feel *love* and *trust* for them, and to believe they can feel these emotions in return. What psychological benchmarks have so far failed to measure, however, is the *authenticity of the relationship,* a quality that is being increasingly devalued. Turkle proposes a broader view of benchmarks that includes consideration of the *long-term impact* of sociable robots on the individual and on society and of how they are transforming our ideas about what it means to be a person.

We believe the papers of this special issue and last year's symposium will provide a sound basis for the future development of psychological benchmarks for human–robot interaction. These benchmarks will not only gauge and guide progress in robot design but will also deepen our understanding of the human condition.

Acknowledgments

We would like to express appreciation to the editorial board of this special issue for their gracious assistance and insightful comments during the review process: Colin Allen, Nadia Bianchi-Berthouze, Stephen J. Cowley, Cory D. Kidd, Andrea Kleinsmith, Jessica Lindblom, Will Taggart, Andrea Lockerd Thomaz, and Tom Ziemke. In addition, we thank Mike Brady, Tamami Fukushi, Hanne De Jaegher, Takashi Minato, Christian J. Onof, John Paley, and Ayse Pinar Saygin for their detailed reviews of submissions. Thanks are also due to Jacob Faiola for assistance with copyediting. This special issue was partially supported by the National Science Foundation under Grant No. IIS-0325035. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necesserily reflect the views of the National Science Foundation.

Note

1. Proceedings *The 15th International Symposium on Robot and Human Interactive Communication (RO-MAN 2006),* University of Hertfordshire, Hatfield, UK, 6–8 September 2006, IEEE Press, ISBN: 1–4244–0564–5