Designing Socially Intelligent Robots

CYNTHIA BREAZEAL
Media Arts and Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts

This talk addresses advances in entertainment technology as it applies to robots. Rather than limit this presentation to the state of robotics in the entertainment industry per se, we interpret "entertainment" more broadly to encompass applications for robots where long-term appeal is important. Significantly, the emerging market of personal service robots poses the question of how to design robots that successfully play a role in the daily lives of ordinary people. Beyond performing useful tasks, personal robots must be natural and intuitive for the average consumer to interact, communicate, work with as partners, and teach new skills, knowledge, and tasks (Fong et al., 2003). To address these challenges, new areas of inquiry in the field of autonomous robotics are emerging, including Human-Robot Interaction (HRI) and Social Robotics. In this short paper, we argue that social and emotional intelligence will be fundamental to the design of personal service robots (Breazeal, 2002). After all, personal robots should not only be useful to their human users, but ideally people will genuinely enjoy having their robots around.

HISTORICAL PRECURSORS

The idea of creating life-like robots has amused and fascinated us for thousands of years. Throughout history, humans have sought to mimic the appearance, functionality, and longevity, as well as the cognitive and adaptive processes of biological creatures. As far back as the ancient Greeks, the idea of life-like machines appears in Homer's Iliad where Hephaistos, the god of metal smiths, fashions mechanical helpers—strong, vocal, and intelligent maidens of gold. The idea surfaces again in medieval times in the Jewish legend of the Golem, a robot-like servant made of clay brought to life by Rabbi Loew of Prague.

As technology advanced, people began to actually build such machines. The first technological breakthrough occurred in the 15th century with the ability to build mechanical clocks. One hundred years later, clockmakers extended their craft to build mechanical animals. There is even some evidence that as early as c. 1478, a young Leonardo Da Vinci, conceptualized a humanoid automaton, controllable by a very crude but programmable analog computer composed of cogs and pulleys (Rosheim, 2000). Nearly forty years later, in 1515, Leonardo built his famous self-propelled mechanical lion, commissioned by the Medici, that reportedly walked from its place in the room and opened its breast full of lilies, presenting them as a token of friendship from the Medici to Francis I, King of France. In response to the 18th century craze for animated objects, Jacques de Vaucanson created the famous mechanical duck in 1738 that could flap its wings, eat, and digest grain (which still remains a mystery) (Doyon and Liaigre, 1966). Around the 1830s to 1840s, Joseph Faber invented a mechanical talking head, called Euphonia, which an operator could reputedly make speak in several European languages (Lindsay, 1997). These are just a couple of examples of historical mechanical automata; a more complete account can be found in Rosheim (1994).

The year 1946 marks the invention the ENIAC computer, the first large-scale general-purpose electronic digital computer (McCarney, 1999). Just a few years later, in 1950, the famous British mathematician, Alan Turing, wrote a provocative paper called, “Computing Machinery and Intelligence” where he discusses the possibility of building machines that can think and learn, and outlines a test (the “imitation game” later known as the Turing Test) to determine if a machine can think (Turing, 1950). That same year, Grey Walter published his work on building two robotic tortoises out of analog circuitry that could navigate towards a light source and interact with one another in simple ways (Walter, 1950). In the science fiction arena, Isaac Asimov published his famous three laws of robotics (Asimov, 1942). A visionary Walt Disney applied robotic technology to entertainment purposes for their earliest physically animated performers, such as the famous Abraham Lincoln audio-animatronic that debuted at the 1964-65 New York World's Fair.

MODERN APPLICATIONS

Today we see robotic technology applied to diverse entertainment purposes. We are familiar with animatronics in theme parks and the use of sophisticated robotic puppets for special effects in films. Recent advances in low-cost electronics has enabled new commercial applications where the ability for robots to interact with people in an entertaining, engaging, or anthropomorphic manner is an important part of its functionality. For instance, a new generation of robotic toys has emerged—many are inexpensive, but some are rather sophisticated such as Sony's robotic dog, Aibo. Robotic kits for edutainment, such as Lego's Mindstorms, allows kids and adults alike to create their own robotic inventions. Location-based entertainment applications such as museum tour guide robots (Nourbakhsh, 1999) offer not only entertainment value but also provide visitors with information of interest. Health-related applications are being explored, such as robotic pet therapy surrogates that are intended to provide the same health benefits of their living counterparts. Even robots for scientific purposes are starting to take on more socially interactive qualities. For instance, NASA Johnson Space Center's humanoid robot, Robonaut, is envisioned to be a completely autonomous astronaut's assistant that is able to work as a productive and cooperative member of human-robot teams (Bluethman et al., 2003).

What of the science fiction dream of your very own Star Wars R2-D2 or C-3PO—an appealing robotic sidekick that helps you in your daily life? We are starting to see the precursors of such futuristic visions in university and corporate research labs around the world, such as Honda's humanoid robot, ASIMO. Toyota recently announced their Partner Robot Project with the stated goal of developing humanoid robots that function as personal assistants for humans. These robots shall "have human characteristics, such as being agile, warm and kind and also intelligent enough to skillfully operate a variety of devices in the area of personal assistance, care for the elderly, manufacturing and mobility."

Robotic Trends magazine defines personal service robots as "robots or robotic technology purchased by individual consumers that educate, entertain, or assist, or protect in the home." One of the strongest motivating applications for personal robots is to provide the elderly with domestic assistance and care. The global demographic trend of rapidly aging societies, where a smaller working age population is responsible for supporting a larger retired population, is the most urgently pressing application for bringing robots into the homes as capable assistants for people and supplementing the workforce. The IMF predicts that Japan, in particular, will experience a dramatic change in their ratio of working age to retired age people—from 4:1 today to 2:1 by 2025. In addition, the convergence of many technological developments in mobile computing such as advances is microprocessor technology, wireless technology, image processing, speech recognition, motor sensor technology, and embedded systems development tools, are making personal robot development increasingly feasible.

Although the service robot market is immature, the few quantitative studies that do exist indicate that the personal service robot market on the verge of dramatic growth. Recent research by the Japan Robotics Association (JRA), United Nations Economic Commission (UNEC), and the International Federation of Robotics (IFR) indicates that the service robot market will experience exceptional growth both in the near term, from $600M in 2002 to approximately $6B in 2009, and expand even more quickly for the next couple of decades, reaching an estimated $60B by 2025. Of course, one must always take with a significant grain of salt such extrapolations from existing studies to predict the future of immature markets. However, if such predictions come to pass, personal robots will be a ubiquitous technology.

THE PSYCHOLOGY OF ROBOT DESIGN

The success of personal service robots hinges not only on their utility but also on their ability to be responsive to and interact with ordinary people in a natural and intuitive manner. Furthermore, given that personal robots are envisioned to coexist with people on a daily basis, issues of how to design for long-term appeal will impact our willingness to accept them into our lives. For instance, longitudinal studies to assess the successful adoption and impact of assistive technologies for the elderly have shown that functionality and need are only part of the design equation. Social and emotional factors also greatly impact the person’s willingness to adopt the technology. Technologies that are stigmatizing, i.e., make the user feel feeble or vulnerable, or make the user feel that they appear that way to others, are often rejected. Even worse when adopted, however, stigmatizing technologies contribute to self-imposed isolation or depression (Forlizzi, et al., 2004). Designing to support human psychology, beyond cognitive considerations, to include social and emotional factors will be just as important for the design of personal robots.

According to *The Design of Everyday Things*, it is essential for people to have a good conceptual model for how entities operate, whether it is a device, a robot, or even another person, in order to interact successfully with them (Norman, 1990). With such a model, it is possible to explain and predict what the other is about to do, its reasons for doing it, and how to elicit a desired behavior from it. The design of a technological artifact, whether it is a robot, a computer, or a teapot, can help a person form this model by "projecting a image of its operation," either through visual cues or continual feedback. Adhering to natural signals and mappings (e.g., physical metaphors or social norms) makes the artifact intuitively understandable to people.

Numerous HCI studies suggest that people will apply a social model when observing and interacting with autonomous robots (Kiesler and Goetz, 2003). The studies of Reeves and Nass show how people treat even desktop computers as social entities and adhere to social norms in their interactions with them (Reeves and Nass, 1996). In fact, such studies demonstrate that it takes surprising few cues to unleash our human social psychology—even a text interface alone is sufficient. Autonomous robots are quite different from desktop computers in their projected animacy. As with living things, the behavior of autonomous robots is a product of their internal state as well as physical laws. They perceive their world, make decisions, and perform coordinated actions to carry out tasks. Augmenting such self-directed, creature-like behavior with the ability to communicate with, cooperate with, and learn from people further encourages people to anthropomorphize them (even for simple vehicles such as those described in Braitenberg, 1984).

We refer to the class of autonomous robots that are explicitly designed to encourage people to apply a social model to interact with and understand them as *social robots*. Designing social robots with personality may help provide people with a good mental model for them. According to Norman (2004), personality is a powerful design tool for helping people form a conceptual model that channels beliefs, behavior, and intentions in a cohesive and consistent set of behaviors. From a design perspective, the emotion system of a robot could implement much of the style and personality of the robot, encoding and conveying its attitudes and behavioral inclinations toward the events it encounters. The robot's personality must be designed such that its behavior is understandable and predictable to people. Therefore, parameters of the personality must fall within recognizable human (or animal) norms, otherwise the robot may appear mentally ill or completely alien. The science of natural behavior, as well as artistic insights from classical animation and character design (Thomas and Johnson, 1981), are a useful guide in this respect.

ROBOTS WITH SOCIAL AND EMOTIONAL INTELLIGENCE

As robot designers, we tend to emphasize the cognitive aspect of intelligence when designing robot architectures while viewing the social and especially the emotional aspect with skepticism (see Sloman, 1981, for an exception). However, numerous scientific studies continue to reveal the reciprocally interrelated roles that cognition and emotion play in intelligent decision-making, planning, learning, attention, communication, social interaction, memory, and more (see Isen, 2000, for a review). Two conceptually distinct and complementary information processing systems, cognition and emotion, evolved under social and environmental pressures to promote the health and optimal functioning of the creature (Damasio, 1994). As argued by

Norman, Ortony and Russell (2003), the cognitive system is responsible for interpreting and making sense of the world, whereas the emotion system is responsible for evaluating and judging events to assess their overall value with respect to the creature (e.g., positive or negative, desirable or undesirable, etc.).

Emotion plays an important role in signaling the salience of things, to guide attention toward what is important and away from distractions, thereby helping to effectively prioritize concerns (Picard, 1997). Alice Isen has studied the numerous beneficial effects that mild positive affect has on a variety of decision making processes for medical diagnosis tasks (Isen, 2000), e.g., facilitating memory retrieval; promoting creativity and flexibility in problem solving; and improving efficiency, organization and thoroughness in decision making. While negative affect allows us to think in a highly focused way when under negative, high-stress situations, conversely, positive affect allows us to think more creatively and to make broader associations when in a relaxed positive state.

Furthermore, scientists are finding that whereas too much emotion can hinder intelligent thought and behavior, too little emotion is even more problematic. The importance of emotion in intelligent decision-making is markedly demonstrated by Damasio's studies of patients with neurological damage that impairs their emotional systems (Damasio, 1994). Although these patients perform normally on standardized cognitive tasks, they are severely limited in their ability to make rational and intelligent decisions in their daily lives. For instance, they may lose a lot of money in an investment. Whereas healthy people would become more cautious and stop investing in it, these emotionally impaired patients do not. They cannot seem to learn the link between bad feelings and dangerous choices, so they keep making the same bad choices again and again. The same pattern is repeated in their relationships and social interactions resulting in loss of jobs, friends, and more.

By looking at highly functioning autistics, we can see the crucial role that emotion plays in normal relations with others. They seem to understand the emotions of others like a computer—memorizing and following rules to guide their behavior but lacking an intuitive understanding of others. They are socially handicapped, not able to understand or interpret the social cues of others to respond in a socially appropriate manner (Baron-Cohen, 1995).

Such scientific findings provide valuable insights and lessons for the design of autonomous robots that must operate in complex and uncertain environments and perform in cooperation with people. This talk presents a pragmatic view of the role emotion-inspired mechanisms and capabilities could play in the design of autonomous robots as applied to human-robot interaction (HRI). Given our discussion above, many examples could be given to illustrate the variety of roles that social and emotion-inspired mechanisms and abilities could serve a robot that must make decisions in complex and uncertain circumstances, either working alone or with other robots. Our primary interest, however, concerns how social and emotion-inspired mechanisms can *improve* the way robots function in the human environment and their ability to work effectively in partnership with people.

This endeavor does not imply that these emotion-based or cognition-based mechanisms and capabilities must be in some way identical to those in natural systems. In particular, the question of whether or not robots could have and feel human emotions is irrelevant to our purposes. Furthermore, the insights these social-based and emotion-based mechanisms provide robot designers should not be glossed over as merely building "happy" or entertaining robots. To do so is to miss an extremely important point: as with living creatures, these social and emotion-inspired mechanisms modulate the cognitive systems of the robot to make it function *better* in a complex, unpredictable environment—to allow the robot to make *better* decisions, to learn more effectively, to interact more appropriately with others, etc. than it could with its cognitive system alone. Therefore, as we continue to design integrated systems for robots with internal mechanisms that complement and modulate its cognitive capabilities with those regulatory, signaling, biasing, and other useful attention, value assessment, and prioritization mechanisms associated with emotion systems in living creatures, then we will effectively be giving robots a system that serves the same useful functions that emotions serve in us—no matter what we call it.

The purpose of this short paper is to set the stage for why social and emotional intelligence will be important in the design of personal robots that assist and entertain their human users. Our accompanying presentation explores issues related to the design of sociable robots from artistic, scientific, and technological perspectives (Breazeal, 2002). Specific research projects are highlighted in the talk to illustrate how robots with social-emotive capabilities are being applied to assist human astronauts in space, to perform opposite human actors in film, to serve as a learning companion for children, and more.

REFERENCES

Asimov, I. 1942 (Reprinted 1991). I Robot. New York: Bantam Books.

Baron-Cohen, S. 1995. Mindblindness. Cambridge, Mass.: MIT Press.

Bluethmann, W., R. Ambrose, M. Diftler, E. Huber, M. Goza, C. Lovchik, and D. Magrude. 2003. Robonaut: a robot designed to work with humans in space. Autonomous Robots 14(2-3): 179-207.

Braitenberg, V. 1984. Vehicles, Experiments in Synthetic Psychology. Cambridge, Mass.: MIT Press.

Breazeal, C. 2002. Designing Sociable Robots. Cambridge, Mass.: MIT Press.

Damasio, A. 1994. Decartes Error: Emotion, Reason, and the Human Brain. New York: G. P. Putnam's Sons.

Doyon, A., and L. Liaigre. 1966. Jacques Vaucanson, mécanicien de genie (Jacques Vaucanson, Genius Mechanic). Paris: Presses Universitaires de France.

Fong, T., I. Nourbakhsh, K. Dautenhahn. 2003. A survey of socially interactive robots. Robotics and Autonomous Systems 42(3-4): 143-166.

Forlizzi, J., C. Di Salvo, F. Gemperle. 2004. Assistive robotics and an ecology of elders living independently in their homes. Human-Computer Interaction 19: 25-59.

Isen, A. 2000. Positive affect in decision making, edited by M. Lewis and J. Haviland. The Handbook of Emotions, Second Edition. New York: The Guildford Press.

Kiesler, S., and J. Goetz. 2002. Mental models of robotic assistants. Pp. 576-577 of the Proceedings of CHI 2002 Conference on Human Factors in Computing Systems. New York: ACM Press.

Lindsay, D. 1997. Talking head. American Heritage of Invention & Technology 13(1): 57-63.

McCartney, S. 1999. ENIAC: The Triumphs and Tragedies of the World's First Computer. New York: Walker & Company.

Norman, D., A. Ortony, and D. Russell. 2003. Affect and machine design: lessons from the development of autonomous machines. IBM Systems Journal 41(1): 39-44.

Norman, D. 1990. The Design of Everyday Things. New York: Basic Books.

Norman, D. 2004. Emotional Design. New York: Basic Books.

Nourbakhsh, I., J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. 1999. An affective mobile robot with a full time job. Artificial Intelligence 114(1-2): 95-124.

Picard, R. 1997. Affective Computation. Cambridge, Mass.: MIT Press.

Reeves, B., and C. Nass. 1996. The Media Equation. Palo Alto, Calif.: CSLI Publications.

Rosheim, M. 1994. Robot Evolution: The Development of Anthrobotics. New York: Wiley.

Rosheim, M. 2000. L'automa programmabile di Leonardo." XL Lettura Vinciana. 15 aprile 2000. Citta' di Vinci. Biblioteca Comunale Leonardiana. Giunti Gruppo Editoriale, Firenze.

Sloman, A., and M. Croucher. 1980. Why robots will have emotions. Pp. 197-202 of the Proceedings of the Seventh International Conference on Artificial Intelligence. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence

Thomas, F., and O. Johnson. 1981. The Illusion of Life. New York: Hyperion.

Turing, A.M. 1950. Computing machinery and intelligence. Mind 59(236): 433-460.

UNECE and IFR (United Nations Economic Commission for Europe and The International Federation of Robotics). World Robotics 2002. New York and Geneva: United Nations Publications.

Walter, W.G. 1950. An imitation of life. Scientific American 182: 42-54.