Abstract—This paper explores the topic of human–robot interaction (HRI) from the perspective of designing sociable autonomous robots—robots designed to interact with people in a human-like way. There are a growing number of applications for robots that people can engage as capable creatures or as partners rather than tools, yet little is understood about how to best design robots that interact with people in this way. The related field of human-computer interaction (HCI) offers important insights, however autonomous robots are a very different technology from desktop computers. In this paper, we look at the field of HRI from an HCI perspective, pointing out important similarities yet significant differences that may ultimately make HRI a distinct area of inquiry. One outcome of this discussion is that it is important to view the design and evaluation problem from the robot’s perspective as well as that of the human. Taken as a whole, this paper provides a framework with which to design and evaluate sociable robots from a HRI perspective.

Index Terms—Human–robot interaction (HRI), socially guided learning, social or sociable robot partner.

I. INTRODUCTION

HUMAN–ROBOT INTERACTION (HRI) is a newly emerging field that has been gaining an increasing amount of interest by researchers in the field of autonomous robotics, as well as those in human-computer interaction (HCI). Traditionally, autonomous robots have been targeted for applications requiring very little (if any) interaction with humans, such as, sweeping minefields, inspecting oil wells, search and rescue, or, exploring other planets. Such robots are viewed as sophisticated tools that are directed remotely by a human supervisor. Service robot applications, such as delivering hospital meals, mowing lawns, or vacuuming floors, bring autonomous robots into environments shared with people [1], but traditionally HRI in these tasks is still minimal—people being more often treated as obstacles to be navigated around, rather than as social beings with which to cooperate.

However, recent commercial applications are emerging where the ability to interact with people in an entertaining, engaging, or seamless manner is an important part of the robot’s functionality. A new generation of robotic toys has emerged (such as Tiger Electronic’s hamsters-like Furby or Sony’s robotic dog, Aibo) whose behavior changes the more children play with it. Although the ability of these products to interact with people is limited, they are motivating the development of increasingly life-like and socially sophisticated robots. Projects, such as Aurora, are exploring the use of robots to play a therapeutic role in helping children with autism [2]. Location-based entertainment applications, such as museum tour guide robots [3], offer not only entertainment value, but also provide visitors with information of interest.

Mediated communication through robotic avatars is another potential application (i.e., extending teleconferencing to robotic conferencing). Here, the robotic manifestation allows one to have a physically embodied and social presence to others—allowing all to share the same reference frame and facilitating the ability to use deictic gestures, to make eye contact, to greet another by shaking their hand, etc.).

Other applications include “wearable” robots, such as robotic exoskeletons to help enhance the physical abilities of the elderly, or robotic prosthetics that replace a lost ability of a disabled person.

Corporate and university research labs are exploring application areas for robots that assist people in a number of ways. Here, the robot is viewed more as a collaborator, assistant, or pet rather than as a tool. For instance, Robonaut is a humanoid robot under development at the National Aeronautics and Space Administration’s (NASA) Johnson Space Center to ultimately serve as an astronaut’s assistant. NEC corporation is developing a small mobile household robot (called PaPeRo) to help people interact with electronic devices around the house (e.g., TV, computer, answering service, etc.). Health-related applications are also being explored, such as the use of robots as nursemaids to help the elderly [4], or robotic pets (such as Omron’s NeCoRo) that are intended to provide some of the health related benefits of pet ownership. The commercial success of these robots hinges not only on their utility, but also on their ability to be responsive to and interact with people in a natural and intuitive manner.

II. PARADIGMS OF HRI

From these numerous examples and applications, one can classify the field of HRI into four interaction paradigms. These are the following:

- robot as tool;
- robot as cyborg extension;
- robot as avatar;
- robot as sociable partner.

Each is distinguished from the others based on the mental model a human has of the robot when interacting with it. In the first paradigm, the human views the robot as a tool that is used to perform a task. The amount of robot autonomy varies (and hence, the cognitive load placed on the human operator) from complete teleoperation, to a highly self-sufficient system that need only be supervised at the task level. In the second paradigm, the robot is physically merged with the human to the extent that the person accepts it as an integral part of their body.
For instance, the person would view the removal of their robotic leg as an amputation that leaves them only partially whole. In the third paradigm, the person projects him/herself through the robot in order to communicate with another from far away—the next best thing to being there. The robot provides a sense of physical presence to the person communicating through it, and a sense of social presence to those interacting with it. The last paradigm speaks to the classic science-fiction fantasy of an artificial being. Interacting with it is like interacting with another socially responsive creature that cooperates with us as a partner.

Although each of these paradigms sounds quite distinct from the others, there are a few shared challenges. First, in each case there is an aspect of shared control between robot and human. For instance, an autonomous explorer is capable of self-navigation. A cyborg extension might have basic reflexes (such as quickly withdrawing from intense heat) to avoid damage, or require tight local feedback from its synthetic skin to grasp a fragile object without breaking it. A robot avatar needs to coordinate speech, gesture, gaze, and facial expression, and direct them to the correct person at the right time. Finally, a robot partner shares control of the dialog and the exchange of speaking turns with its human interlocutor.

The ability to effectively share control gives rise to another important issue, i.e., the ability to appropriately understand the intention (or internal state) of the other. This is important for both parties in order to coordinate and synchronize their behavior. It allows them to work effectively as a team, to correct misunderstandings before success is compromised, and to compensate for unexpected difficulties before failure becomes manifest. For instance, to carry out a particular task, a human supervised robot must know which goal its operator wants it to accomplish. Conversely, to monitor the robot, the human needs to understand what the robot is trying to do. At a lower level of interaction, a robot prosthetic hand needs to know when to pick up an object, and the person needs to know when the grip is secure. A robot avatar needs to understand the intent behind a given message to convey it appropriately through gesture or facial expression (i.e., is the person being sarcastic, humorous, or serious), and the person needs some feedback that what he/she meant was communicated appropriately. During social interaction with a robot partner, both parties need to appropriately convey their intended meaning to the other and assess if it was received appropriately.

III. HCI APPLIED TO SOCIABLE ROBOTS

All of these areas of HRI are important and fascinating areas of research. This paper focuses on the last paradigm, robot as sociable partner [5]. As these kinds of robots take on an increasingly ubiquitous role in society, they must be easy for the average person to use and interact with. This raises the important question of how to properly interface untrained humans with these sophisticated technologies in a manner that is intuitive, efficient, and enjoyable to use. In the field of HCI, Reeves and Nass [6] have shown that humans (whether computer experts, lay-people, or computer critics) generally treat computers as they might treat other people provided that the technology behaves in a socially competent manner. From their numerous studies, Reeves and Nass argue that a social interface may be a truly universal interface given that humans have evolved to be experts in social interaction.

From these findings, we take as a working assumption that attempts to foster human–robot relationships will be accepted by a majority of people if the robot displays rich social behavior. Similarity of morphology and sensing modalities makes humanoid robots one form of technology particularly well-suited to this. If the findings of Reeves and Nass hold true for sociable robots, then those that participate in rich human-style social exchange with their users offer a number of advantages. First, people would find working with them more enjoyable and would thus, feel more competent. Second, communicating with them would not require any additional training since humans are already experts in social interaction. Third, if the robot could engage in various forms of social learning (imitation, emulation, tutelage, etc.), it would be easier for the user to teach new tasks. Ideally, the user could teach the robot just as one would teach another person.

While robotics researchers tackle the technical issues of building autonomous robots for these new human-centered applications, these efforts could benefit from the techniques and methodologies of the HCI community in evaluating human–robot interaction (HRI). Various task domains need to be explored including functional scenarios where robots might help a person perform a physical task, educational scenarios where a robot might help in adult training or participate in educational games for children, health scenarios where a robot might provide assistance to the elderly or disabled, or entertainment scenarios where the goal is a rewarding and compelling interaction.

HCI-like studies as applied to HRI could be used to advance a scientific understanding of how people interact with this type of robotic technology. This, in turn, would inform how to engineer robots that interact effectively with people. Design issues include the robot’s morphology (e.g., should it be more anthropomorphic, creature-like or vehicle-like?), aesthetic appearance (e.g., should it appear organic or mechanical?), physical skillfulness, perceptual capabilities, communicative expressiveness, and its intelligence (e.g., social, emotional, or cognitive). Such design issues would be well served by HRI studies that addressed the following issues:

1) Comparative Media Issues: How does interacting with robotic technologies differ from other interactive media (such as software agents)? In what ways is it similar? Are there special affordances that a robotic media offers that could be leveraged from in order to improve HRI? How might this compare to mixed-media applications such as merging robotics with graphical animation?

2) Naturalness Issues: How are people naturally inclined to interact with this sort of technology? In what ways will people try to teach it? This impacts the kinds of interaction scenarios that the robot’s design must support. Will they engage it as they would another person (using natural social cues, etc.). If not, then in what ways might this differ?

3) User Expectation Issues: What are people’s implicit expectations for the robot’s capabilities? For instance, do people expect the robot to communicate using natural language? Do
they expect the robot to understand what they are feeling? How can you design the robot to shape or calibrate the person’s expectations to be commensurate with the robot’s capabilities? This can mitigate the person’s disappointment or frustration when interacting with the robot. It can also gently steer the person to interact with the robot in the way it was intended.

4) Quality Issues: How does one design robots that are enjoyable, useful, and rewarding for people to interact with? What aspects make the robot more appealing and engaging? What aspects make the robot intimidating or annoying?

5) Relationship Issues: What should be the nature of the human–robot relationship? Should it be more like interacting with a tool/appliance, a creature/pet, or a person (e.g., collaborator/supervisor/servant)? What social roles are appropriate for robots?

6) Teamwork Issues: How can robots serve as effective members of a human–robot teams? Clearly robots must be designed so that they are competent at their tasks. They must also be able to effectively communicate and cooperate with people. Teamwork issues also arise, such as how to integrate robots into teams so that the human members accept them (e.g., training with people, etc.), utilize them to the best of their ability, and trust them appropriately to get the job done.

7) Personality Issues: How does the person’s personality impact the design of the robot? Should the robot be designed to convey a personality itself? If so, of what type and how complex?

8) Cultural Issues: How do cultural attitudes impact the design? Many communicative styles, gestures, and mannerisms are culture specific. Those that might be considered polite or friendly in one culture might be rude in another (such as personal space, the use of touch, when to make eye contact or to avert gaze). What kinds of behavior are socially acceptable verses inappropriate for a robot? Social structures (and where robots might fit within them) vary between cultures, dictating a robot’s mannerisms (e.g., its degree of formality).

9) Acceptance Issues: Science fiction has promoted a favorable view of robots in Japanese society, whereas it has contributed to a more suspicious viewpoint in American culture. How will this impact the way in which robots are accepted and integrated into human culture? How does this impact attitudes toward what robots should do, should not do, or cannot do? How accountable are robots for their actions?

IV. Different Kind of Technology

HCI has much to offer with respect to designing technologies that support human needs. Does this imply that HRI is simply an adaptation of HCI to robots?

Although the challenge of building autonomous robots that interact with people may share some issues with the design of computer interfaces, robots and computers are profoundly different technologies in important ways. In this section, we highlight these differences with respect to long term autonomy in the real world, the ability to interact with people, and the ability to learn from people.

Based on these key differences, we argue that it is not sufficient in HRI to evaluate a robot’s behavior solely based on the human’s perspective (as is the case in HCI). It is important to recognize that both robot and human are part of a system, and it is the performance of the human–robot system that ultimately matters. Both members have goals that relate to the task at hand, and both have extenuating circumstances that they must tend to (e.g., the need to survive, the need for self-maintenance, the ability to take advantage of a learning opportunity, etc.). If well designed, the relationship can be mutually beneficial—each can help the other and each can learn from the other. Therefore, it is important to examine and evaluate matters from the robot’s point of view as well!

A. Long-Term Interaction

A robot is part of the physical environment—it shares our world with us. It is likely that an owner would encounter his/her robot on a daily basis, either by intentionally seeking it out, by chance encounters as as the robot goes about performing its chores, or perhaps initiated by the robot seeking out the person. It is not quite the same with software agents where a person must go to their computer (or look at their PDA, or open their cell phone, etc.) to interact with it. In other words, there are times when people choose to interact with the world of information, and times when they do not. In contrast, people must always deal with the physical world. The opportunity for frequent interaction over an extended period of time (potentially for years), and the opportunity to establish a long-term relationship, poses some significant design challenges for robots.

B. Survival in the Real World

Robots not only have to carry out their tasks, they also have to survive in the human environment. The ability for robots to adapt and learn in their environment is fundamental given that human designers cannot predict all possible circumstances and challenges a robot will encounter during its lifetime (unless the task and environment are very structured). Human society is a particularly challenging environment given its richness, its dynamic nature, its unpredictability, and its uncertainty (imagine the complexity of everyday family life in the home to a robot). It is an environment that is not easily simplified without imposing significant restrictions which might be unacceptable to the people that share that environment. Nonetheless, robots must perform tasks and make decisions given imperfect and partial knowledge and information. Hence, much of robotic design addresses issues of robustness, adaptivity, and dealing with uncertainty—all in addition to the specific knowledge and skills required to perform a certain task. In contrast, software agents tend to deal with more specialized tasks in more restricted environments.

C. Deeply Integrated “Interface” and “Control”

The computer model of having a clean division of the “interface program” from the underlying “application program” is not easily made with autonomous robots. The “interface” is not a layer that sits at the surface, producing the robot’s observable behavior that mediates the interaction between the human and the underlying control mechanisms that carry out the task. Rather, it is the observable behavior that allows the robot to ne-
gotiate its way about the real world—whether it is physically manipulating objects, socially engaging people, or dealing with self-maintenance functions. It is quite possible that the robot uses one to do the other and vice versa (e.g., asking a person to open a door for it so that it can dock with its recharging station inside). Hence the functionality of the robot cannot be easily partitioned into “interface” behavior, “task” behavior, and “survival” behavior. The social and emotive qualities of a robot serve not only to “lubricate” the interface between itself and its human interlocutor, but also play a pragmatic role in promoting survival, self-maintenance, learning, decision-making, attention, and more [7]–[9].

D. Interacting With People

When interacting with a human, sociable robots bring an interesting set of affordances. Certainly, some of these affordances are shared with other interactive media, such as embodied conversational agents [10]. For instance both can perceive the naturally offered social cues of a human using cameras or microphones. These might include perceiving the person’s tone of voice, articulated speech, facial expression, articulated gesture, body posture, and so forth. Furthermore, both have bodies (either animated or mechanical) to deliver these same social cues to a person.

To different degrees, both can share the same reference frame with a human. This is useful for exchanging deictic gestures or for establishing a shared referent through gaze direction and/or head pose. However, this is clearly more limited for a character restricted to a screen with statically mounted sensors, than for a robot whose sensors can move with it. Similarly, it is more difficult for an animated character to establish and maintain compelling eye contact given the limitations of a planar screen. Humans are exquisitely perceptive of gaze direction and eye contact, and we have found that this ability has powerful impact on a person’s sense of being engaged on a personal and direct level [11].

There are other affordances that seem particular to having a physical embodiment. For instance, robots have the ability to manipulate real objects to perform physical tasks. They are also able to locomote and move in the same physical space as people. There is the possibility for direct physical contact between robots and people, such as shaking a person’s hand in a greeting. A human might touch it or physically interact with a robot as a pet. This introduces interesting benefits as well as possible risks. A technology is not so easily dismissed when it has the ability to proactively seek you out and come into immediate contact with you.

E. Learning in the Human Environment

As stated above, beyond communication and interaction, any robot that co-exists with people as part of their daily lives must be able to learn and adapt to new experiences. Ultimately, people will be able to teach the robot how to do new tasks, or the particulars of how to do a given task. For instance, even a task as specific as taking out the trash has a number of distinct variables, such as locating a particular trash can in a specific home, opening that style of trash can, navigating through that home and yard, scheduling when to remove the trash, and so forth.

Hence, one key challenge is to design robots that are as easy to teach as another person. Ideally the robot could engage in various forms of social learning such as imitation, emulation, tute-lage, etc. Today, humanoid robots (or physics based simulations of them) can learn a specific physical skill by observing a human demonstration of it [12], can acquire a simple proto-language by engaging in imitative interactions with a human instructor [13], or can mimic a sequence of human gestures by learning a mapping from the human’s body to their own [14], [15].

V. Socially Guided Learning

Although such work has dominantly focused on articulated motor coordination, there are many advantages that social cues and skills could offer robots that learn from people. A socially competent robot could take advantage of the same sorts of social learning and teaching scenarios that humans readily use. Below are five key challenges of robot learning, and how social, emotional, and expressive factors can be used to address them in interesting ways.

A. Knowing What Matters

Faced with an incoming stream of sensory data, a robot must figure out which of its myriad perceptions are relevant to learning the task. As its perceptual abilities increase, the search space becomes enormous. If the robot could narrow in on those few relevant perceptions, the learning problem would become significantly more manageable.

Knowing what matters when learning a task is fundamentally a problem of determining saliency, which can be guided either internally or externally [16]. Objects can gain saliency because of their inherent properties (motion, color, proximity, etc.). Objects can also become salient if they are the focus of the instructor’s attention as indicated through gaze direction, language, or deictic gestures. Such social and guiding cues also help the learner to identify the most relevant items to consider. This guidance accelerates state-space discovery, where the machine learns new groups of features that have behavioral significance.

To facilitate this process, the state of the learner’s attention must be transparent to the instructor so that he or she can easily infer what the learner is attending to, and what it is about to do. This state information can be conveyed through familiar social cues such as gaze direction, head orientation, and body pose [11]. Roy and Pentland [17] have relied on this approach to have a robot learn the semantics of words.

B. Knowing What Action to Try

Once the robot has identified salient aspects of the scene, how does it determine what actions it should take? As robots become more complex, their repertoire of possible actions increases. This also contributes to a large search space. If the learner had a way of focusing on those potentially successful actions, the learning problem would be simplified.

Determining which action to try can be addressed in a number of ways. The robot could experiment on its own by selecting an action based on past experience. However, a human instructor can play an important facilitating role in guiding the learner’s
exploration of the most promising actions. If the learner already knows how to perform the associated action, then a person might simply tell it what to do. However, sometimes the agent will have to learn how to perform the action if it is not already present in its repertoire. In this case, a human instructor could provide considerable assistance by demonstrating the appropriate actions to try—especially if the human and robot share the same morphology as shown in [18]. Alternatively, this action-space discovery would be facilitated if it is easy to guide the system into performing desired action such as using traditional animal training techniques such as shaping or luring, or through mimicry or imitation, as demonstrated in [19].

C. Knowing When to Learn and Who to Learn From

When should a robot either exploit what it already knows or explore new possibilities? Knowing when to explore relates to how predictable and/or controllable the world is for the robot. Adding a reflective aspect could help a system explore more insightfully on its own—or at the very least, know when it needs help to explore, and when it needs to find an appropriate teacher to help guide that exploration.

Exhibiting expressions of inquisitiveness in the presence of human teachers can also assist in this process of knowing when to learn, or learning when to learn. Humans can choose when to reward or encourage such curious behavior from the machine, and when to redirect the machine’s learning toward more relevant topics. We expect that the social-emotional skills of communication between the machine learner and the human teacher will become especially important in such frequent interactions. It will be important for the machine to curry favor from the human teacher and not to irritate him or her. Thus, machine will receive a greater amount of attention and guidance, which will aid its goal to learn. In our past work, we have demonstrated that systems that display childlike expressions of inquisitiveness successfully elicit teaching behaviors from adults and from children [5].

D. Correcting Errors and Recognizing Success

Once a learner can observe an action and attempt to perform it, how can the robot determine whether it has been successful? How does it assign credit for that success? Further, if the learner has been unsuccessful, how does it determine which parts of its performance were inadequate?

This requires a reflective ability for the robot to assess its own learning progress. It must be able to identify the desired outcome and to judge how its performance compares to that outcome. In many situations, this evaluation depends on understanding the goals and intentions of the instructor as well as the agent’s own internal state. Additionally, the agent must be able to diagnose its errors in order to improve performance.

Fortunately, the human instructor has a good understanding of the task and knows how to evaluate the learner’s success and progress. One way that a human instructor could facilitate the learner’s evaluation process (to recognize success and correct failures) is by providing feedback through a number of channels. Facial expression, gesture, speech, tone of voice, etc. all provide feedback that allows the learner to determine progress and whether it has achieved the goal.

In this way, the human can play an important role in guiding the exploration of the robot through intuitive communication channels [5]. To support this process, it must be easy for the instructor to tell what the agent has learned and what it has not learned yet. The agent must also be able to communicate to the human what it is sure about and what it is confused about. It must assign credit in a way that matches the trainer’s expectation. The robot must be a transparent learner.

E. Leverage From Provided Structure

Finally, the instructor can use the learner’s expressions as feedback to control the rate of information exchange—to either speed it up, to slow it down, or to elaborate as appropriate [20]. By regulating the interaction in partnership with the learner, the instructor can establish an appropriate learning environment and provide better quality instruction.

The ability to take turns lends significant structure to the learning episode that the learner can use to incrementally refine its performance. The instructor demonstrates, the learner performs, and then the instructor demonstrates again, often exaggerating or focusing on aspects of the task that were not performed successfully. To support this, the robot’s observable behavior must change in a way that provides feedback to the instructor, and in a way that motivates the instructor to teach.

VI. CONCLUSION

Taking this body of work as a whole, we argue that endowing a robot with social skills and capabilities has benefits far beyond the interface value for the person who interacts with it. The ability for robots to interact with people and to leverage from these interactions to perform tasks better, to promote their self-maintenance, and to learn in an environment as complex as that of humans is of tremendous pragmatic and functional importance for the robot.

The desire to bring autonomous robots into the social world of people poses challenges beyond traditional applications of remote operations. To survive and function in our world, we evolved and developed social intelligence and emotional intelligence. Given this, it is perhaps not so surprising that introducing robots into the same environment may require that we endow them with forms of socio-emotional intelligence for the same reasons (see our accompanying paper [21] in this volume). To be useful for the robot to this extent, such characteristics cannot be restricted to the surface (at the “interface”), but integrated deep into the core of their design.

Their performance and the benefits they bring to us will still need to be evaluated, of course, but from the human’s perspective and that of the robot. Developing such dual measures for autonomous sociable robots may make HRI a related, yet distinct area of inquiry from HCI. In some cases, a more ethological based methodology may be needed to accommodate more free-form interactions between humans and robots. Clearly a strong dialog is needed between the robotics, HCI, and other related communities in order to establish appropriate techniques, measures, studies, etc. It is our hope that this paper lends some insight into the nature of this work, and offers a step toward what the field of HRI will become.
ACKNOWLEDGMENT

The author would like to acknowledge that this paper is the result of the ongoing efforts of the graduate and undergraduate students of the MIT Media Lab’s Robotic Life Group and our collaborators.

REFERENCES


Cynthia Breazeal received the B.S. degree in electrical and computer engineering from the University of California, Santa Barbara, and the M.S. and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1993 and 2000, respectively. She is an Assistant Professor of Media Arts and Sciences at MIT. Her interests focus on human-like robots that can interact, cooperate, and learn in natural, social ways with humans.