WHAT'S THE BEST ROLE FOR A ROBOT?
Cybernetic Models of Existing and Proposed Human-Robot Interaction Structures

Victoria Groom
Department of Communication, Stanford University, 450 Serra Mall, Stanford, U.S.A.
vgroom@stanford.edu

Keywords: Cybernetics, human-robot interaction, robot teams, social robotics.

Abstract: Robots intended for human-robot interaction are currently designed to fill simple roles, such as task completer or tool. The design emphasis remains on the robot and not the interaction, as designers have failed to recognize the influence of robots on human behavior. Cybernetic models are used to critique existing models and provide revised models of interaction that delineate the paths of social feedback generated by the robot. Proposed robot roles are modeled and evaluated. Features that need to be developed for robots to succeed in these roles are identified and the challenges of developing these features are discussed.

1 INTRODUCTION

Human-robot interaction (HRI) is the study of humans' interactions with robots. While the field of robotics focuses primarily on the technological development of robots, HRI focuses not just on the robot, but on the broader experience of a single or group of humans interacting with robots. Researchers have long sought to deploy robots alongside humans as human-like partners, minimizing humans' involvement in dangerous or dull tasks. While robots have demonstrated some promise as coordination partners, in practice they contribute little to achieving humans' goals, often requiring more attention and maintenance and eliciting more frustration than their contributions are worth. Through these failures, it has become clear that not only must robots' technical abilities be improved; so must their abilities to interact with humans.

Humans prefer that all interaction partners that exhibit social identity cues display role-specific, socially-appropriate behavior (Nass & Brave, 2005; Reeves & Nass, 1996). A robot must cater to this human need to facilitate a successful interaction, but designers of robots are rarely attuned to human psychological processes.

Discounting human needs and expectations has led HRI researchers to propose design goals for robots that fail to fully consider the needs of humans. Creating a "robot teammate" has become a guiding goal of the HRI community, even though the needs and expectations of humans intended to team with robots have not been properly considered (Groom & Nass, 2007). Because HRI has yet to become a fully-established field, putting careful thought into the goals of HRI now is essential for its future success.

Cybernetics--the study of complex systems, particularly those that feature self-regulation--places a strong emphasis on the value of modelling interactions and provides an established framework for understanding and talking about systems, something much needed in HRI. While HRI researchers often model systems within a robot, little attention has been paid to modelling the interaction between a human and a robot.

Cybernetic models featuring a goal, comparator, actuator, and sensor clearly delineate the relationship between systems and their environments. The system's goal is to affect the environment in some manner within some parameters. The system's comparator determines if the goal has been achieved and transmits this information to the actuator, which takes some action on the environment. A sensor then detects some feature of the environment, and this information is passed to the comparator. With cybernetic models, systems continually influence and are influenced by their environments and other systems.

In this paper, I draw on cybernetics to represent the models shaping the design of robots intended for close human interaction. I critique these models and offer revised models that include the human, the robot, and the interaction between them. I also model the conversational abilities required of teammates, identifying those features that must be developed in robots for humans to accept them. The difficulty of meeting these requirements raises questions as to whether the field of HRI is pursuing optimal goals.

2 HRI DESIGN TODAY

Today's robots are not yet capable of serving in roles like teammate that require sophisticated social capabilities. While designers are working on creating robots capable of filling these roles, the majority of existing robots fill less demanding roles. These roles have lower requirements for autonomy, intentional action, and socially-appropriate behavior, and are similar to those roles filled by other advanced technologies such as computers.

2.1 Robot Roles

One role that robots are often designed to fill is *task completer*. In this role, robots complete a task designated by a human. Many military robots, such as bomb-detecting and bomb-defusing robots, are modelled in this role. In some cases the robot's system may be non-cybernetic and in others it may be cybernetic. With non-cybernetic task-completer robots, the human sets the goal of the robot and the robot affects the environment in a manner intended to achieve the goal. In the case of a bomb-detecting robot, the robot may run tests on a potential bomb and send data back to distantly-located humans. The process terminates at this point, as the system lacks a sensor, comparator, or both. The process used by the human to select the goal is not modelled, nor is there any indication that the robot's behavior affects the humans' goals.

A cybernetic task completer is generally more robust and capable of more complex tasks than a non-cybernetic task completer. The Roomba is a popular example of a cybernetic task-completer robot. As with non-cybernetic task completers, the goal of a cybernetic task completer is set by a human. Unlike non-cybernetic task completers, the Roomba features sensors and a comparator that partly comprise a cybernetic system, which enables the Roomba to navigate obstacles. As with the non-cybernetic task-completer, the human is considered only peripherally in the design process. In the case of the Roomba, the human is modelled as having little interaction with the robot. The human provides the robot power, maintains and cleans it, and initiates its activities by turning it on.

The *tool* is another model commonly used for the design of robots. A tool extends humans' influence on the environment or grants humans power over the environment that they do not normally possess. Because a robot tool is much like an extension of the self, attention is paid to the human operating the robot: the goals and processes of the humans are often considered in the design of the robot. The robot is designed to help a human complete a task or range of tasks. As a tool, the robot is outside the human system, acting within the environment on the environment.

Search and rescue robots often take the form of a tool. One reason robot tools are useful in search and rescue situations is because they enable people to examine and influence areas that are inaccessible or too dangerous for humans to access (Casper & Murphy, 2003). The model of the robot tool differs from models of the robot task-completer in that the influence of the robot on the human is acknowledged. However, the influence of the robot is indirect, as the human senses only the environment which contains the robot. Additionally, the influence of the robot on the human is limited to the humans' selection of the best means to implement a task strategy. The design of the robot as tool does not model the robot as influencing the human directly nor directly affecting the human's higher level goals, such as selecting a task strategy.

2.2 Social Feedback

The existing models of robots as task completors and tools fail to delineate the powerful direct influence of the robot on the human. Most designers of robots, even those within the HRI community, fail to fully recognize the social feedback that robots generate. The behaviors of humans that interact with bomb-detecting and defusing robots, Roomba, or search and rescue robots indicate that they are receiving information from the robot beyond that which is intentionally designed.

An ethnographic study of the use of the Roomba in family homes found that half of all families studied developed social relationships with it (Forlizzi & DiSalvo, 2006). These families named the robot, spoke to it, described social relationships between it and pets, and even arranged "play dates" for multiple Roombas to clean together. In addition, the Roomba affected the cleaning strategies of household members, with males assuming a greater involvement in house-cleaning. Anecdotal evidence suggests that soldiers who interact closely with robots in high stakes situations, like bomb diffusion and search and rescue, form close emotional bonds with robots, giving them names and grieving when the robots sustains serious injuries.

In these cases, humans are responding to social information generated by the robot. Computers as Social Actors theory (CASA) was developed by Nass (Reeves & Nass, 1996). CASA posits that that even when technologies lack explicit social cues, people respond to them as social entities. Research performed under this paradigm has shown that even computer experts are polite to computers (Nass, Moon, & Carney, 1999), apply gender stereotypes to computers (Lee, Nass, & Brave, 2000), and are motivated by feelings of moral obligation toward computers (Fogg & Nass, 1997). Even unintentional cues of social identity elicit powerful attitudinal and behavioral responses from humans.

Research indicates that some of the reasons that people respond to computers socially is because computers exhibit key human characteristics (Nass, Steuer, Henriksen, & Dryer, 1994), including using natural language (Turkle, 1984) and interacting in real time (Rafaeli, 1990). Robots generally demonstrate even more human characteristics than robots. Some robots, such as Asimo or Robosapien, feature a humanoid form. Many robots, such as Nursebot or Roomba, feature some form of locomotion, an indicator of agency. In addition, robots often exhibit at least some autonomous action and appear to humans to sense their environments, make judgments, and act on their environments. The very nature of robots make them appear even more like social entities than most other existing technologies and elicit an even more powerful social response. But only when one of the primary design goals is to foster a social relationship, as with entertainment robots like Aibo or Robosapien, is the social influence on the robot considered.

As indicated in Figure 1, the robot's behavior has a powerful influence on operators' higher-level goals. People have a high-level goal of recognizing and evaluating social information transmitted by others. Humans wish to respond to the behavior of others in a socially-appropriate manner (Reeves & Nass, 1996). While designers may have intended for the robot to be an invisible tool, it is in fact sending powerful cues indicating that it is a social entity. The robots behavior may affect humans' task strategies, either through direct feedback or by influencing humans' higher level goals to act socially appropriately.

Recognizing the influence of social feedback on

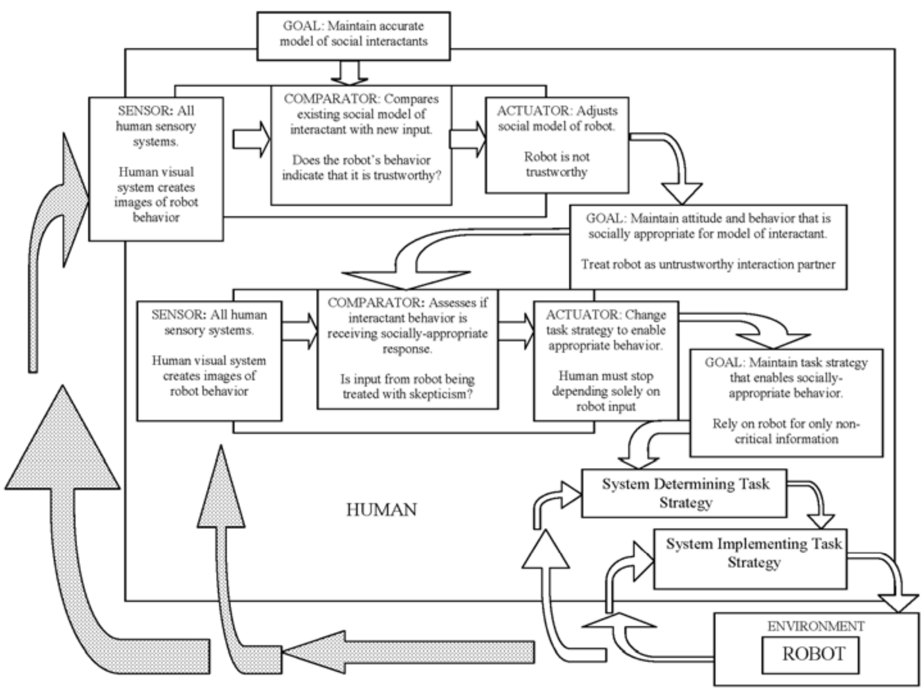

Figure 1: Model of a robot tool. Social feedback indicated with dotted arrows. Component boxes contain examples.

humans interacting with robots has important implications for the design of robots. Designers are more likely to consider which aspects of their design are likely to generate a social response from humans. Designers may be more inclined to create intentional cues to foster a social relationship or to elicit the desired social response. For example, it has been demonstrated that humans apply gender stereotypes to voices—even those that are obviously synthetic (Nass & Brave, 2005). Awareness of this effect may lead designers to choose robot voices not only based on the clarity of the robot's voice, but also based on the desired social response.

Considering social feedback when designing robots plays a key role in setting humans' expectations of robots. The fewer and weaker the cues of social identity, the lesser the likelihood is that a robot will elicit a social response. Robotocist Masahito Mori (1970) coined the term the "Uncanny Valley" to describe humans' responses of discomfort when a robot's visual or behavioral realism becomes so great that humans' expectations of human-like behavior are set too high for the robot to meet. When a robot is less realistic, humans have lower expectations and are able to tolerate non-humanlike behavior. As visual and behavioral indicators of humanness increase and human-like behaviour doesn't, people become negative. Only when the humanness of robots' behaviors catch up to their highly human-like appearance will robots emerge from the valley of uncanniness. When designing robots for interaction with humans, recognizing the role of social information in setting user expectations will enable designers to manage social cues and set expectations that the robot is capable of satisfying.

3 FUTURE OF HRI DESIGN

The roles that robots are successfully filling today, such as task completer and tool, fail to take advantage of robots' full potential. Computers also succeed in these roles, but robots have features that computers do not. Robots have the potential to move about their environments, sensing the world around them, and either transmitting that information to distantly-located humans or making decisions and acting on the environment directly.

The ultimate goal for designers involved with HRI is to create a robot capable of serving as a member of a human team. Few researchers have sought to define "team" or "team member" or identify the requirements for creating a robot team member. The robot team member has been generally accepted as a lofty but worthy and attainable goal. (For a summary and criticism of the "robot as teammate" model, see Groom & Nass, 2007).

A well-established body of research is dedicated to the study of teams. Successful teammates must share a common goal (Cohen & Levesque, 1991), share mental models (Bettenhausen, 1991), subjugate individual needs for group needs (Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004), view interdependence as positive (Gully, Incalcaterra, Joshi, & Beaubien, 2002), know and fulfill their roles (Hackman, 1987), and trust each other (Jones & George, 1998). If a human or robot does not meet these requirements, they may never be accepted into a team or may be rejected from the team when problems arise (Jones & George, 1998).

One key requirement of teammates that underlies all other requirements is the ability to engage in conversation with other teammates. To be a successful conversation partner, a robot teammate must be able to both convey meaning in a way that other teammates can understand and understand the meaning intended in the communications of other teammates. If a robot cannot do this, human teammates can never be certain if the team shares a common goal, which makes the human unable to trust the robot in risky situations. Likewise, humans would be uncertain if the robot was subjugating its needs, viewing interdependence as positive and knowing and fulfilling its role. Without conversation, humans would feel certain that the robot was incapable of sharing a mental model.

Figure 2 provides a model of conversation between teammates that is derived from cybernetician Gordon Pask's (1975) Conversation Theory (CT). One key element of this model of conversation is the emphasis on *both* conversation partners' involvement in the communication. Another related element is that both partners construct the meaning of a message in their mind. Meaning is not directly transmitted from one conversation partner to the other, so each partner must be capable of deriving meaning from a message. A successful conversation requires that each person not only ascribe their own meaning to messages, but also infer the meaning of others and compare the meaning of each partner to determine if they are in agreement. While some robots are capable of recognizing words or gestures and responding appropriately, no robot has come close to being able to fully engage in conversation.

Figure 2 highlights those features that must be developed in robots for them to achieve the most basic requirement of teammates: the ability to engage in conversation. These requirements may be broken down into three general categories: concepts, knowledge, and systems. To communicate and behave in a manner that allows humans to interpret meaning, robots must demonstrate awareness of basic concepts, including goals and motivation. Robots lack humans' complex hierarchy of goals. Human teammates deployed in a high-stakes situation like search and rescue maintain many goals at once, including a goal to survive, a goal to protect other teammates, and a goal to succeed at the task at hand. Robots maintain a limited number of simple goals that are always set at some point by a human.

One of the most important areas of knowledge that robots lack is an understanding of common human motivating factors and the relationship between specific goals, motivations, and actions. In order for robots to be useful in uncontrolled, changing situations, they must posses a broad body of knowledge. Robots' lack of knowledge of common goals, motivations, and actions also make them difficult for humans to understand, eliciting unintended negative responses from humans. While a human's motivation to avoid harm encourages void, and destroy itself. Human teammates are likely to feel frustration, disappointment, and betrayal when a robot acts in a manner that is self-destructive and detrimental to the team.

In order for robots to be accepted by humans in situations that rely on conversations and mutual dependence, robots must exhibit behavior that appears to humans to imply an underlying systems much like the system used by humans to create and use mental models. Human conversation partners rely on their own mental models and their abilities to create mental models of others' mental models. While it is possible that robots could successfully fake mental models, they must rely on a system that can serve a similar purpose to mental models and appear to humans as a mental model. If robots are unable to do this, humans will never feel certain they share the same model of a goal.

Teams rely on a willingness of teammates to subjugate their own personal goals for a team goal. To do this, robots must demonstrate a sophisticated goal hierarchy and effective communication skills. Teams also depend on a high level of trust. Any breakdowns in conversations may result in the unraveling of the team. Maintaining trust requires

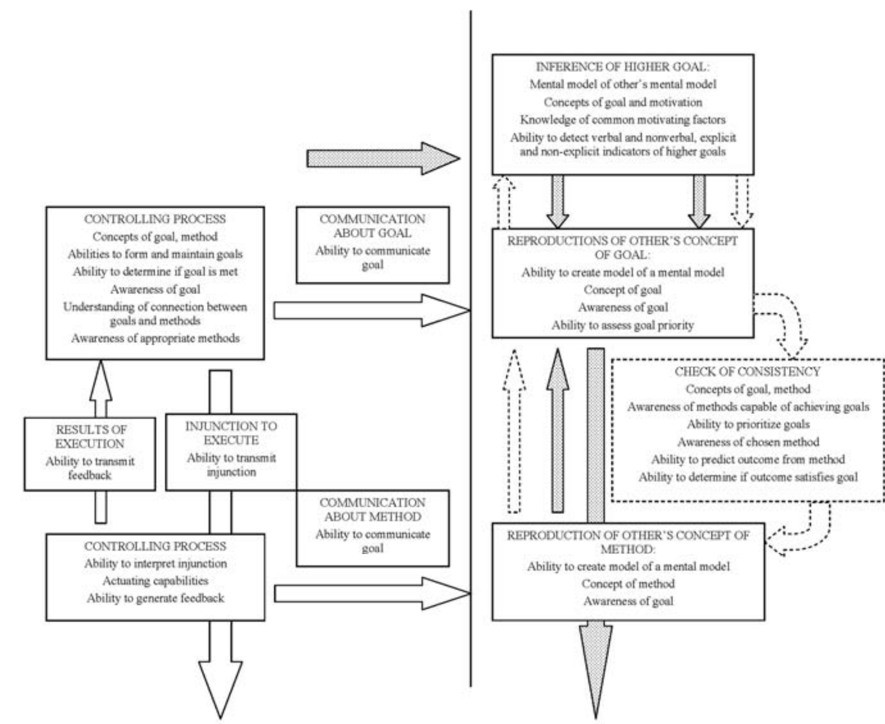

Figure 2: Model of a robot conversation partner. Robot must be capable to both communicate (left side) and interpret (right side) meaning. Robot abilities that must be developed are indicated in each box. Dotted arrows indicate inferences. Dashed arrows indicate checks of consistency.

and seek to repair the relationships (Jones & George, 1998). Even if a robot meets the basic requirements of a conversation partner, its conversational abilities will need to be further developed to meet the higher expectations of teammates.

4 CONCLUSIONS

If designers wish to place robots in roles that have previously been filled only by humans, they must design robots that demonstrate the social behavior and communication skills that humans expect of people in these roles. To create robot teammates, robots' concepts of goals, motivations, actions, and the relations between them must become further developed and nuanced. Achieving this requires the development of systems so complex that they generate behaviors that enable humans to infer the existence of shared mental models. Once researchers recognize that creating a robot teammate takes far more than improving a robot's performance and introducing it into a human team, the HRI community can weigh the challenges of developing a robot teammate to determine if creating a robot teammate is indeed the best goal to guide the direction of HRI research.

ACKNOWLEDGEMENTS

The author thanks Hugh Dubberly, Paul Pangaro, and Clifford Nass for their feedback and support.

REFERENCES

Bettenhausen, K. L. (1991). Five years of groups research: What we have learned and what needs to be addressed. Journal of Management, 17(2), 345-381.

Casper, J., & Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. Systems, Man and Cybernetics, Part B, IEEE Transactions on, 33(3), 367-385.

Cohen, P. R., & Levesque, H. J. (1991). Teamwork. Nous, 25(4), 487-512.

Fogg, B. J., & Nass, C. (1997). Do users reciprocate to computers? Paper presented at the ACM CHI.

Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment: a study of the roomba vacuum in the home. ACM SIGCHI/SIGART Human-Robot Interaction, 258-265.

Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human–robot teams. Psychological Benchmarks of Human–Robot Interaction: Special issue of Interaction Studies, 8(3), 483-500.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. Journal of Applied Psychology, 87(5), 819-832.

Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), Handbook of organizational behavior. Englewood Cliffs, NJ: Prentice Hall.

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: implications for cooperation and teamwork. The Academy of Management Review, 23(3), 531-546.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intelligent Systems 19 (6), 91-95.

Lee, E.-J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotypes. Paper presented at the CHI 2000, The Hague, The Netherlands.

Mori, M. (1970). The uncanny valley. Energy, 7(4), 33-35.

Nass, C., & Brave, S. B. (2005). Wired for speech: How voice activates and enhances the human-computer relationship. Cambridge, MA: MIT Press.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. Journal of Applied Social Psychology, 29(5), 1093-1110.

Nass, C., Steuer, J., Henriksen, L., & Dryer, D. C. (1994). Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to "self-" or "other-" evaluations. International Journal of Human-Computer Studies, 40(3), 543-559.

Pask, G. (1975). An Approach to Machine Intelligence. In N. Negroponte (Ed.), Soft Architecture Machines: MIT Press.

Rafaeli, S. (1990). Interacting with media: Para-social interaction and real interaction. In B. D. Rubin & L. Lievrouw (Eds.), Mediation, Information, and Communication: Information and Behavior (Vol. 3, pp. 125-181). New Brunswick, NJ: Transaction.

Reeves, B., & Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people and places: Cambridge University Press New York, NY, USA.

Turkle, S. (1984). The second self: Computers and the human spirit. New York: Simon & Schuster.