Can robots be teammates?

Benchmarks in human–robot teams

Victoria Groom and Clifford Nass
Stanford University

The team has become a popular model to organize joint human–robot behavior. Robot teammates are designed with high-levels of autonomy and well-developed coordination skills to aid humans in unpredictable environments. In this paper, we challenge the assumption that robots will succeed as teammates alongside humans. Drawing from the literature on human teams, we evaluate robots' potential to meet the requirements of successful teammates. We argue that lacking humanlike mental models and a sense of self, robots may prove untrustworthy and will be rejected from human teams. Benchmarks for evaluating human–robot teams are included, as are guidelines for defining alternative structures for human–robot groups.

Keywords: human–robot interaction, psychological benchmarks, robot teams, social robotics, teamwork, trust

Identifying the best model for human–robot interaction

There are many social structures a small group of people can adopt to achieve a goal, including hierarchy, autonomous individuals relying on "an invisible hand," pure divisions of labor, and cliques. However, *teams* have proven to be the most powerful and effective structure (S. G. Cohen & Bailey, 1997). Membership in a team enhances communication, trust, effort, and commitment (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990; Back, 1951). Membership in a team also reduces common social distinctions such as race and gender, as teammates focus on similarities rather than differences (Allen & Wilder, 1975, 1979; Mackie & Smith, 1998; Sebanz, Bekkering, & Knoblich, 2006)

Teams are not only effective; they are also seductive. Research under the minimal group paradigm demonstrates that given the smallest excuse to form and feel part of a team, people will leap at the opportunity (Tajfel, 1982). Even something

Interaction Studies 8:3 (2007), **483–500**.
ISSN 1572–0373 / E-ISSN 1572–0381

as trivial as arbitrarily labeling people as members of the "green team" as opposed to the "orange team" induces a sense of identity, mutual dependence, and positive perceptions of other "greens" (Turner, 1982, 1985). An emphasis on similarities, especially shared goals, also can induce the sense that a person is part of a "team" and its concomitant positive outcomes. Indeed, one could argue that humans are "built" to be part of a team (Nass, Fogg, & Moon, 1996; Reeves & Nass, 1996).

Teams are not universally successful: other species find different approaches to be more effective. Geese that fly in the classic "vee" formation, for example, are jockeying for position (weak birds are pushed to the outside) in an "every bird for itself" methodology that is more efficient than an attempt to support team membership through communication (Dawkins, 1989). Not even all social animals rely on teams: Ants, for example, seem to have little sense of other ants in the colony (Wilson & Holldobler, 1990). The number of animals that exhibit anything close to "team" behavior is remarkably small (Dawkins, 1989). Instead, each species finds the means to achieve its goals via the social structure that is best suited to the species' strengths and weaknesses (Wilson, 2000).

Human teams are so effective that there is a natural tendency to enlist nonhumans as teammates. What happens when nonhuman animals are brought into human teams? While there is an expectation that human teams are so powerful and effective that they should be able to gracefully incorporate at least one or two animal others, especially given the power of teams to blur differences between members, the results have generally been poor, even when the proposed teammates are social animals such as dogs. In the best-studied examples, there have been attempts to bring dogs into human search and rescue teams as full-fledged teammates. No matter how much the dog trains with the team as a whole, it has proven to be much more effective to provide the dog with a single "handler" who is part of the team: the dog is an adjunct to the handler rather than a team member in its own right.

Before the advent of robots, there was a general understanding that technologies could not be teammates. Noninformation technologies have long been understood as at most facilitators of teams. Although people may appreciate the capabilities of computers, few see them as having the potential to function as teammates (Nass, Fogg, & Moon, 1996). In essence, computers do not seem to have the required communication and coordination abilities.

Although animals and technologies have not even been pretenders for team membership, the team has been designated the normative organizational model for human–robot interaction. The 2007 International Conference of Human–Robot Interaction was themed "Robot as Team Member." In the official program, the Co-Chairs note that robots used in "critical domains… must coordinate their behaviors with human team members; they are more than mere tools but rather

quasi-team members whose tasks have to be integrated with those of humans." It is not suprising that people address the need for coordination by modeling robots on teammates. Teams appear so valuable that it may seem they should be considered first among other possible social structures. Unlike animals, robots can be designed specifically to support human activities. Robots can be programmed to focus unwaveringly on the team's goal. Researchers continue to make significant strides in training robots to engage in socially-appropriate behavior and coordinated action (Fong, Nourbakhsh, & Dautenhahn, 2003). There is also evidence that robots and especially androids elicit a broad range of social responses to a greater extent than computers and other technologies (MacDorman & Ishiguro, 2006a, 2006b). Given how humanlike robots can appear, the answer to the question, "can and should we try to design robot teammates?" has been "yes," and the field has moved on to the question, "how do we make robots *better* teammates?"

In this paper, we problematize the notion that robots can and should be teammates. We argue that the benefits of teams to human organization may have led researchers and designers astray in attempting to tune robots for human-motivated teams. Instead, we argue that the strengths and weaknesses of robots require a re-thinking of how one should attempt to optimize coordinated action between humans and robots.

The assumption of "humanness"

Human teams are successful because they take advantage of the strengths of humans. For example, all people have "mental models." A mental model is a conceptual framework for describing, explaining, and predicting experience (Rouse, Cannon-Bowers, & Salas, 1992). Humans' mental models are similar enough and flexible enough for a team to share (within some degree of approximation) one team mental model that guides joint activity. Sharing one mental model enables a team to make more efficient decisions than decisions made by individual teammates (Klimoski & Mohammed, 1994). Because teams leverage skills and processes that come naturally to humans, the optimized features are difficult to identify. It is not obvious that teams take advantage of the structure of *human* mental models, because mental models are so fundamental to human cognition.

Researchers thus have not noticed the extent to which this human quality is assumed of teammates. The assumption that the team is an ideal model for human–robot interaction is based on the even more basic assumption that the best model for the development of robots is the human model. When researchers ask the question, "How do we make robots better teammates in human teams?" what they are really asking is, "How do we make robots better *human* teammates?" The view that robots can succeed as teammates while computers cannot may de-

rive from the fact that robots are ascribed a greater potential to exhibit characteristically human behavior.

By presupposing that robots could be teammates, researchers may miss noticing the possibility that imbuing robots with the "humanness" assumed of teammates may be extremely challenging or even impossible. If robots cannot or will not manifest "humanness," then the question, "Is it possible to make robots successful teammates?" needs to be reevaluated. If robots can never be effective teammates, the team model should be replaced with a more appropriate model.

The human ability to accept robots into teams ultimately will determine the success or failure of robot teammates. Although the technical capabilities of robots to engage in coordinated activity is improving (as visible for robot-only teams at RoboCup[1]), we believe humans' innate expectations for team-appropriate behavior pose challenges to the development of mixed teams that cannot be fixed with technological innovation. Since we focus on the human response to robots positioned as teammates, we limit our discussion to human–robot teams in which robots and humans work together as teammates in one team. Our discussion of robot teammates does not apply to partnerships, teams of humans interacting with robots external to the teams, or teams of robots controlled by external operators. "Teammate" has a strict organizational definition, which we describe in detail below.

Human–robot organizational structures

Murphy (2004) divides robot applications into two broad domains: robots deployed in controlled environments and robots deployed in unpredictable environments. The latter collection of applications she terms *field applications*; these include space exploration, SWAT teams, military robotics, and rescue robotics. Teams prove particularly appealing in these areas because the autonomy, flexibility, and coordination of teammates strengthen the group's ability to adjust to unforeseen circumstances and changing situations. In field applications, operators are usually physically distant from the robots they control and breakdowns in communication may have negative consequences. High stakes make explicit the need for a robust coordination model, such as the team. Robots in field applications are often operated by a human team (Murphy, 2004); extending the team to include the robot appears a logical adaptation.

If a robot is not identified as a teammate, it may be treated as a tool. The robot-as-tool model, often used in teleoperation, constrains the robot's performance to the operator's skill and the design of the interface (Fong, Thorpe, & Baur, 2002). While the team model promises the robot the autonomy needed to liberate it from humans' restrictive control, we argue that robots are currently unable to meet

humans' high expectations for appropriate team behavior in the unpredictable, high-stakes situations for which they are designed.

Our problematization of the human–robot team is not a criticism of other human–robot interaction structures. In controlled environments, robots in human-like social roles are demonstrating great promise. The public's positive response to commercial entertainment robots, such as Sony's Aibo and WowWee's Robosapien, highlights the promise of robots as entertainers. As robots become increasingly adept at playing games with humans, the effectiveness of robot entertainers will likely increase (Brooks et al., 2004). The Huggable and other similar haptic devices demonstrate the ability of robots to serve as media, conveying human communication across distances (Stiehl et al., 2005). Users have responded enthusiastically to iRobot's Roomba, suggesting that domestic servant robots will be welcomed into homes (Forlizzi & DiSalvo, 2006). The utility of robots as therapeutic aids has been particularly well-demonstrated. Both as companions (Libin & Libin, 2004; Tamura et al., 2004; Turkle, 2007) and behavioral therapists (Feil-Seifer, Skinner & Matarić, 2007; Werry, Dautenhahn, Ogden, & Harwin, 2001), robots have proven their ability to enhance people's quality of life. In controlled environments, such as a therapy session, or when limited autonomy is required, such as with entertainment, robots fit easily into human roles. While robots may not be fully convincing, their behavior is rarely in such blatant violation of human expectations as to break the social contract of the role.

Features of successful teammates

To determine if robot teammates can succeed without full-blown humanness, researchers should first identify those qualities that make a successful human teammate. A substantial body of literature on human teams identifies benchmarks that are relevant to a team's success. Broadly speaking, teammates need to share interests (Rouse, Cannon-Bowers, & Salas, 1992) and engage in pro-team behavior (Mead, 1934). More specifically, successful teammates must

- Share a common goal (P. R. 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)
- Trust each other (G. R. Jones & George, 1998)

Teams adopt shared goals because forming teams to achieve shared goals makes it easier to achieve individual goals. Teams establish and maintain goals by sharing a team mental model. A team mental model has a form similar to that of an individual's mental model. While each teammate's mental model is unique, they are structurally similar. Each teammate may infer the basic motivations, perceptions, and vulnerabilities of other teammates from her own model. The ability to infer others' mental models enables teammates to develop a team mental model complete with a shared set of goals, strategies, and motivations. Sharing a mental model aids decision-making (Walsh & Fahey, 1986), communication, and collaborative action (H. H. Clark, 1996).

Before engaging in activities to achieve shared goals, potential teammates need to deduce that the benefits of participation in the team outweigh the anticipated personal sacrifices. To make these sacrifices, teammates have to value the group's interests above their own and subjugate personal needs to the needs of the group. Comparing individual needs to the needs of the group requires a sense of self, the ability to distinguish self from other, and personal needs. Furthermore, teammates must view the interdependence of the group as positive. If a person feels that pursuing a goal as a group is less likely to produce the desired reward than pursuing it alone, the individual will not make the sacrifices demanded of teammates (Cannon-Bowers, Salas, & Converse, 1993).

Communication between teammates enables the team to distribute work efficiently (Bales, 1951). If teammates do not know their roles and do not fulfill their duties, the team will perform less well. Similarly, if a team member's actions do not meet the behavioral expectations of other teammates, the strength of the team will deteriorate (Butler, 1976).

The importance of trust

Trust is the feeling of confidence that another individual will not put the self at risk unnecessarily (Anderson, 1980; Axelrod, 1976). Believing that teammates will protect each other's interests enables individual teammates to surrender their personal interests to the interests of the group. Trust between teammates is essential for the successful functioning of a team, as trust bridges shared interests and pro-team behavior. Trust also establishes behavioral expectations that facilitate joint activity (Mayer, Davis, & Schoorman, 1995). In high-risk situations, trust assumes even greater importance (Das & Teng, 2004). Because robots are often designed to take the place of humans in high-risk situations, people's trust of robots will be particularly critical if robots are to succeed as teammates.

Willingness to surrender the protection of one's safety depends in large part on the degree to which individuals trust each other. At-risk teammates must believe

that the team's interest in their welfare is comparable to their own (Mayer, Davis, & Schoorman, 1995) and think that other teammates are capable of protecting their interests (Deutsh, 1960).

The specter of the human–robot team

No group of humans and robots has met the requirements of a team as described above. One reason the team has been accepted as the ideal model for human–robot interaction may be that researchers are unaware of its specific requirements, and therefore underestimate the challenges of creating a team. Researchers often label humans and robots interacting together as teams, but examining these "teams" offers more evidence for the enormity of the challenge of creating true teams than evidence of their successful implementation.

When researchers report the successful implementation of a human–robot team, these "teams" are in fact one of the other successful human–robot organizational structures, such as the partnership. For example, (Sierhuis et al., 2003) developed a human–robot team model that positioned robots as nonessential servants to human needs aboard the International Space Station. This organizational structure would not have been called a team if a subset of cohabitating humans was dedicated entirely to serving the needs of other members. Without personal needs, expendable robots are unable to willingly subjugate individual needs for the benefit of the group, to trust or be trusted, or to view interdependence as positive. In this model, robots are servants and humans are masters.

Unlike humans, robots have not yet exhibited the abilities required of teammates. They do not have unique viewpoints or mental models and cannot be trusted in a humanlike way. Robots lack values, and they cannot access the mental model shared by human teammates. Robots do not exhibit a drive for self-preservation and lack the ability to link motivations to protect the self with motivations to protect the group. Simply put, robots fail as teammates. They set false expectations of autonomy, agency, and self-preservation, making collaboration difficult.

These problems are exacerbated by the human tendency to respond socially to nonhuman stimuli (MacDorman & Ishiguro, 2006a). Humans project a social identity on media and technologies that offer the slightest social cue (Reeves & Nass, 1996). In the case of robots — even nonhumanoids — implicit social cues are powerful. Characterizing robots as teammates indicates that robots are capable of fulfilling a human role and encourages humans to treat robots as human teammates. When expectations go unmet, a negative response is unavoidable.

Attempts at human–robot teams

While there have been few attempts to deploy robots as meaningful teammates in real-life situations, many researchers are focusing on developing in robots the potential to one day succeed in true human–robot teams. Groups at NASA have developed robots intended to take the place of humans in dangerous situations in space exploration and to support humans in the Space Station (Fong, Nourbakhsh, Kunz, Fluckiger, & Schreiner, 2005; Sierhuis et al., 2003). The Stanford Aerospace Robotics Laboratory has worked with the Palo Alto-Mountain View Regional SWAT Team and MLB Company to integrate an Unmanned Aerial Vehicle, the MLB Bat, into a human SWAT team. The MLB Bat provided SWAT teammates on the ground with aerial views of the site. Robin Murphy and others at the University of South Florida have focused on deploying human–robot teams to aid with search and rescue operations. Other groups, such as The Naval Research Laboratory and the Robotics Institute at Carnegie Mellon, have focused on the technical development of features required for robots to act as teammates. Work by these groups include the development of dialog systems (Rybski, Yoon, Stolarz, & Veloso, 2007), modeling of cognitive and affective systems (Cassimatis, Trafton, Bugajska, & Schultz, 2004; Gockley, Simmons, & Forlizzi, 2006), improvement of spatial awareness (Trafton et al., 2005; Trafton et al., 2006), implementation of flexible autonomy structures (Heger & Singh, 2006), and robot training (Rybski, Yoon, Stolarz, & Veloso, 2007).

While these groups are succeeding in improving robots' performance, they have not yet addressed some of the features most essential for human–robot team success, nor have they demonstrated that currently identified challenges can be overcome. For example, the Peer-to-Peer Human Interaction Project was created in response to the NASA Vision for Space Exploration, which called for the development of a program for space exploration that balanced contributions from both robots and humans (Fong, Nourbakhsh, Kunz, Fluckiger, & Schreiner, 2005). Researchers from NASA, the Robotics Institute at Carnegie Mellon, the National Institute of Standards and Technology, the Naval Research Laboratory, and the Massachusetts Institute of Technology collaborated in one of the largest coordinated efforts to develop successful human–robot teams. The group's three major efforts included the development of the "Human–Robot Interaction Operating System" to facilitate task-oriented dialog exchange, cognitive architectures designed for humans and robots to understand each other, and metrics to evaluate team success. A simulated team of astronauts and multiple robots, including the humanoid Robonaut, worked collaboratively on construction tasks. Performance analysis revealed deficiencies in both human–robot and human–human communication. The robots' inabilities to monitor and communicate their status resulted in the

greatest identified deficiency: time lost in communication between astronauts to determine the robots' status.

The Peer-to-Peer simulations lacked ecological validity, preventing researchers from identifying perhaps the greatest deficiency of the project: the absence of human stress. While the few real-world deployments of humans and robots in field environments identify human stress as a major factor affecting human–robot interaction (Casper & Murphy, 2003), simulations like the Peer-to-Peer Human Interaction Project fail to simulate the psychological experience of human teammates in the field. In low-stakes simulations, failures of the robot have little personal impact, while in space, humans must subjugate the most basic individual goal — survival — for successful team performance. When there is not a major risk, individuals can better tolerate failures of communication and need not worry whether a robot is a full-fledged team member. However, when people must count on robots for their very lives, considerations of whether the robot is a committed "team member" come to the fore. The robot's lack of self-awareness in these instances leads to distrust and a rejection of the robot as a teammate. The current approach to evaluating human–robot teams cannot identify this process; thus, they are insufficient for benchmarking.

The experiences of the few groups dedicated to developing true human–robot teams offer enough information to identify specific challenges to human–robot teams. Finding solutions to some of these problems appears inevitable. For example, human teammates often refer to objects by points of reference, and the objects and points of reference change frequently throughout the interaction. While robots currently struggle in this area, researchers are succeeding in improving their abilities to identify objects by reference point (Fong, Nourbakhsh, Kunz, Fluckiger, & Schreiner, 2005). Though it may be many years before robots can consistently understand human references to objects, there appears to be no major block that would prevent eventual success.

Two major problem areas that show less promise of resolution are the robot's inability to earn trust and lack of self awareness. In high-stress situations, robot operators may experience physical and cognitive fatigue. In extreme cases, operators may forget to bring a robot to or from a disaster site (Casper & Murphy, 2003). Unlike human teammates who think not only about how to do their jobs, but also how to be prepared to do their jobs, robots require an external activation of their autonomy. They must be set-up to perform, but in cases of extreme stress, human teammates may lose their ability to babysit the robot. In these cases, the robot will not be in the appropriate position to complete its task.

In the examples considered, the robots have provided insufficient information on their internal state to operators, so operators have needed to spend much of their time diagnosing a robot's problems. In the wake of the World Trade Center

Disaster, an estimated 54% of time spent in voids was wasted diagnosing problems (Casper & Murphy, 2003). Robots lack pain and may continue to engage in destructive behaviors until they are damaged or destroyed. The difficulties regarding robots' self-awareness stem from their lack of a self-other distinction and a drive for self-preservation. Lacking unconditional autonomy, current robots are never fully self-motivated and engage in behaviors that are not only self-destructive, but damaging to the team.

Perhaps the most daunting problem for the development of human–robot teams is overcoming the challenge of establishing trust between humans and robots. Not surprisingly, trusting something that is unable to trust and to feel guilt or betrayal proves difficult. The less a user trusts a robot, the sooner she will intervene in its progress towards the completion of a task (Olsen & Goodrich, 2003). If she believes the robot's actions contradict a higher-level goal, she will override the robot's completion of a low-level task (Olsen & Goodrich, 2003).

Human teams are easily created and expanded. In crises, teammates generally feel certain of other teammates' high-level goals. Even when new teams are integrated into a larger team, teammates may rely on their understanding of others' goals and mental models to predict the new teammates' behavior. Because robots lack fully-formed mental models and operate using task-specific models of their goals and environments, human teams feel they cannot rely on the robot to act only in ways that protect the team's safety. At the World Trade Center site, robots were sent into the rubble last, following humans, search tools, and then dogs (Casper & Murphy, 2003). In SWAT incidents, certainty and control are highly valued; technologies that show any potential to introduce uncertainty are seen as threats. SWAT teammates may reject robots as sources of risk (H. L. Jones, Rock, Burns, & Morris, 2002).

Benchmarking an ongoing successful team

Throughout the lifespan of a team, the integrity of the team will be challenged. Violations of trust are one of the most serious threats, and with robots, these challenges are among the most difficult to overcome. Without self-interest and humanlike mental models, the introduction of a robot into a human team makes violations of trust and the ensuing consequences highly likely. A successful strategy of response to threats is marked by a series of benchmarks. The best way to understand these benchmarks is to understand how they might fail.

Benchmark 1: Conditional trust is established. In low stakes situations, humans grant potential teammates the benefit of the doubt and show initial conditional trust. In human–robot teams, if the initial interaction with the robot is in a low-risk situation, humans may begin trusting the robot conditionally. If the stakes

are initially high, trust may never be established and the humans will not treat the robot as a teammate. If an initially low stakes situation becomes more risky or potentially rewarding, or if the team's goal changes, the human teammates' trust of the robot will be challenged.

Benchmark 2: Trust is challenged. The greater a human teammate's personal investment in the team's ultimate goal, the lower her threshold for violations of expectations. If the robot works with significant autonomy in an uncontrolled environment, it will inevitably act in an unexpected manner. The robot may fail to complete a routine task, or it may adapt to a new situation in an unexpected way. In high stakes situations, even if the robot's actions do not put humans at direct risk, unpredictability will be seen as a threat to the safety of the group. While reciprocation of expectations causes an upward spiral of team success, unmet expectations initiate a downward spiral in trust. The human's conditional trust and initial positive affect towards the robot will shift to distrust and negative affect (G. R. Jones & George, 1998).

Benchmark 3: Violated party attempts to repair trust. If a teammate whose trust has been violated believes the offender still shares the same values, the trust relationship may be reparable. The violation will signal to the human a need to reevaluate the relationship. She will compare the robot's values to her own, and consider her attitudes and emotions with regards to the robot. If her trust has been violated but not destroyed, she will indicate the violation of trust via outbursts (Frijda, 1988). Thus, early responses to minor trust violations by robots will result in verbal and physical displays of negative affect toward the robot and negative emotion. The human may direct frustration directly at the robot or, given the social inappropriateness of scolding a robot, the human may channel it through alternative paths. She may direct outbursts at other human teammates or simply withhold expression of the violation.

Benchmark 4: Renegotiated relationship is established. While minor violations of trust need not destroy a trust relationship, the offender must respond positively to the wronged party's outburst and change their behavior in accordance with the trust agreement. The offender also needs to make the wronged party's feelings more positive (G. R. Jones & George, 1998). Robots are unlikely to change in accordance with the human's request. Over time, the human will criticize the robot for acting unpredictably, regardless of the action or the expected behavior. Negative outbursts will increase, and as the robot fails to renegotiate the trust relationship, the relationship will spiral downward.

Benchmark 5: Team performance changes. If trust is reestablished, the overall performance of the team will show improvement. Without the shared goals and close social relationship of trust, the human will lack the willingness or desire to assist the robot (M. S. Clark & Mills, 1979). Without assistance, the robot may

fail to complete its tasks. A hallmark of a successful team is a willingness among teammates to perform tasks outside their roles that further the goals of the team. Without shared goals, human teammates will limit the scope of their roles. By completing only the tasks dictated by her role, a teammate may jeopardize the safety of the team. Any dependence on the robot will make the human feel uncomfortable. She will be unlikely to engage in help-seeking behavior, which may delay or prevent the resolution of problems (Walster, Walster, Berscheid, & Austin, 1978). Without a timely resolution of problems, the success of the entire team is jeopardized. Uncertain that all teammates will subjugate their individual needs for the group's goal, the humans may shirk their duties (Holmstrom, 1979).

Communication amongst teammates will be damaged as well. Because the individual is concerned that information will be used to her detriment or to exert power over her, she will withhold information from all distrusted parties (Fama & Jensen, 1983), including the robot and those working closely with the robot. When the trust relationship between human and robot deteriorates beyond repair, the human will lobby to remove the robot from the team. If the human succeeds, a productive human-only team may be established. If the robot is not removed from the team, the human may leave the team.

Benchmarking a new interaction model

If researchers accept that robots may never succeed as teammates, does that mean that human–robot collaborations will always be inferior to human–human collaborations? Should we limit robots to the role of socially unaware tool? To answer these questions, we return to the assumption that the team is the best model for human–robot collaboration. While teams are ideal organizations for humans, they are not ideal for the coordination of all entities. Mixed teams featuring humans and some entity lacking the features optimized by teams are certain to fare poorly in comparison to human teams.

The success of the team model for collaborative human behavior demonstrates that organizational structures should take advantage of the abilities of each entity in the group. Instead of steadfastly maintaining the team model of human–robot interaction and trying to make robots comparable to human teammates, researchers should instead develop an organizational structure that exploits both the special abilities of humans and the special abilities of robots. In the struggle to make robots into people, researchers have not fully identified the human characteristics that robots lack nor the robot characteristics that humans lack. While trying to make robots human, researchers have sometimes overlooked what makes robots special.

The advantages of robots

Teams not only take advantage of human strengths; they also work around human weaknesses. Just as it is challenging to recognize the human strengths that teams optimize, so it is challenging to recognize the human weaknesses that teams minimize. Benchmarking and developing desirable abilities in robots that are not restrained by selfhood may enable designers to create robots that succeed in aspects of coordinated activity where people struggle.

A unique point of view, while it offers many advantages, limits a human's perspective. While humans simultaneously process a range of information through a variety of channels, the perspective of the individual is restricted to one physical location. Humans have access only to information that reaches their locations. Robots need not be limited in this way. Robot sensors may be distinct from the robot and placed throughout an environment. Not only does this enable the robot to perceive information unavailable at the robot's physical location, it also enables the robot to perceive itself directly from a third-person perspective. Because self-perceptions are subject-less, obtaining a clear assessment of the status of the self is often more challenging than assessing the status of others. Robots have the potential to perceive themselves both first-personally and third-personally. Even if humans could perceive information through multiple points of view at once, they would be unable to process the information. The concept of the self and the functioning of mental models require that phenomena be experienced through one unique point of view. Robots may be developed not only to perceive information through multiple points of view, but also to process and assimilate this information into one coherent whole.

Humans rely on external sources of information to orient themselves. Humans must use a process of triangulation that relies on external markers to establish their physical locations. This process is limiting in several ways. External orienting references are constantly changing. If a human does not stay alert to these changes, all reference points may be lost and disorientation may occur. Robots need not rely on variable reference points to establish their physical location. Robots designed with the ability to determine their location regardless of context can have navigation abilities superior to those of humans.

Mental models aid human action and team coordination because they make the processing and use of information manageable. For people, perceiving the world directly without the filtration and organization offered by mental models would result in information overload and impairment of intentional action. People refer to simplified models to make decisions, instead of the constantly changing, information-rich world. While mental models do enable people to avoid information overload, they separate the human experience from phenomenal reality. Mental

models do not perfectly represent reality, as information is filtered out, changed to fit the model, and emotionally colored. While mental models enable humans to act with limited information processing, they restrict the information available to make decisions. Without mental models, robots need not filter perceptions nor alter information to enable efficient processing and decision-making. Robots may be designed with the ability to store and process all sensory input. Processing information without mental models also enables robots to evaluate information without the influence of moods or emotions. In high-risk or emotionally-charged situations, a robot's ability to process all information without the influence of attitudes, moods, or emotions, could enable the robot to process information more accurately and make wiser decisions than human counterparts.

Conclusion

We do not wish to offer here a blueprint for a new human–robot organizational structure. Instead, we hope to inspire designers and researchers to identify and develop the potential abilities of robots that humans do not share. To determine the best interaction model for humans and robots, we believe the following questions need to first be addressed. Answering each question sets a benchmark for determining the best model.

- What are the restrictions of humanness that limit performance?
- Which inabilities of humans can be successfully implemented in robots?
- What organizational structure best optimizes both human and robot abilities?
- If an organizational structure familiar to humans is ideal, do robots have the potential to fulfill the social duties of the role, and is developing those abilities worth the effort?

Researchers would not be using robots for side-by-side interaction if they did not believe that robots have assets to contribute that humans do not. Nonetheless, the human tendency to see "humanness" everywhere has led researchers to impose a model of interaction suitable only for human–human interaction. If researchers wish to optimize what is special about robots, they should recognize that robots and humans need to be designed to complement each other in a way that enables optimal social structures to emerge. Rather than expect robots to meet benchmarks for being human, robots should be evaluated in terms of benchmarks that make them effective as complements rather than duplicates. As we move beyond teams to social structures that allow robots to complement humans rather than be ersatz teammates and ersatz people, the true benefits of robots and humans working together will be realized.

Note

1. RoboCup: URL: http://www.robocup.org/, last accessed 17/5/2007

References

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology, 29*(Pt. 2), 97–119.

Allen, V. L., & Wilder, D. A. (1975). Categorization, belief similarity, and intergroup discrimination. *Journal of Personality and Social Psychology, 32* (6), 971–977.

Allen, V. L., & Wilder, D. A. (1979). Group categorization and attribution of belief similarity. *Small Group Behavior, 10* (1), 73–80.

Anderson, J. R. (1981). Concepts, propositions, and schemata: What are the cognitive units? In J. H. Flowers (Ed.), Nebraska symposium on motivation: Cognitive processes (Vol. 28). Lincoln: University of Nebraska Press.

Axelrod, R. M. (1976). *The structure of decision*. Princeton: Princeton University Press.

Back, K. W. (1951). Influence through social communication. *Journal of Abnormal and Social Psychology, 46* (1), 9–23.

Bales, R. F. (1951). *Interaction process analysis: A method for the study of small groups*. Chicago: Addison-Wesley Press.

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.

Brooks, A. G., Gray, J., Hoffman, G., Lockerd, A., Lee, H., & Breazeal, C. (2004). Robot's play: interactive games with sociable machines. *Computers in Entertainment (CIE), 2*(3), 10–10.

Butler, J. K. (1976). Reciprocity of trust between professionals and their secretaries. *Psychological Reports, 53*, 411–416.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 221–246). Hillsdale, NJ: Erlbaum.

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

Cassimatis, N. L., Trafton, J. G., Bugajska, M. D., & Schultz, A. C. (2004). Integrating cognition, perception and action through mental simulation in robots. *Robotics and Autonomous Systems, 49*(1–2), 13–23.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, M. S., & Mills, J. (1979). Interpersonal attraction in exchange and communal relationships. *Journal of Personality and Social Psychology, 37*(1), 12–24.

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

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*(3), 239–290.

Das, T., & Teng, B. (2004). The risk-based view of trust. *Journal of Business and Psychology, 19*(1), 85–116.

Dawkins, R. (1989). *The selfish gene*. Oxford: Oxford University Press.
Deutsh, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations, 13*, 123–140.
Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *Journal of Law and Economics, 26*(2), 301–325.
Feil-Seifer, D., Skinner, K., & Matarić, M. J. (2007). Benchmarks for evaluating socially assistive robotics. *Interaction Studies, 8*(3). (This issue)
Fong, T., Nourbakhsh, I., Kunz, C., Fluckiger, L., & Schreiner, J. (2005). The peer-to-peer human-robot interaction project. Paper presented at the AAAI Space, Long Beach, CA.
Fong, T., Nourbakhsh, I., Kunz, C., Fluckiger, L., & Schreiner, J. (2005). The peer-to-peer human–robot interaction project. *Space*, 2005–6750.
Fong, T., Thorpe, C., & Baur, C. (2002). Robot as partner: vehicle teleoperation with collaborative control. *Proceedings from the 2002 Naval Research Laboratory Workshop on Multi-Robot Systems*. Washington, DC, USA.
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.
Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43* (5), 349–358.
Gockley, R., Simmons, R., & Forlizzi, J. (2006). Modeling affect in socially interactive robots. *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2006),* (pp. 558–563) Hatfield, UK.
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.
Heger, F., & Singh, S. (2006). Sliding autonomy for complex coordinated multi-robot tasks: Analysis and experiments. *Proceedings Robotics: Science and Systems II*, August 16–19, 2006. Philadelphia, Pennsylvania, USA. The MIT Press 2007
Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics, 10*(1), 74–91.
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.
Jones, H. L., Rock, S. M., Burns, D., & Morris, S. (2002). Autonomous robots in swat applications: Research, design, and operations challenges. *Proceedings of the Association of Unmanned Vehicle Systems International (AUVSI) International Conference on Unmanned Vehicles*, Orlando, FL, USA.
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.
Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management,* 20 (2), 403–437.
Libin, A. V., & Libin, E. V. (2004). Person–robot interactions from the robopsychologists' point of view: The robotic psychology and robotherapy approach. *Proceedings of the IEEE, 92*(11), 1789–1803.
MacDorman, K. F. & Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies, 7*(3), 297–337.

MacDorman, K. F. & Ishiguro, H. (2006). Opening Pandora's uncanny box: Reply to commentaries on "The uncanny advantage of using androids in social and cognitive science research." *Interaction Studies, 7*(3), 361–368.

Mackie, D. M., & Smith, E. (1998). Intergroup relations: Insights from a theoretically integrative approach. *Pscyhological review, 105*(3), 499–529.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model or organization trust.. *Academy of Management Review, 3*(20), 709–734.

Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.

Murphy, R. R. (2004). Human–robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 34*(2), 138–153.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human–Computer Studies, 45*(6), 669–678.

Olsen, D. R., & Goodrich, M. A. (2003). Metrics for evaluating human–robot interactions. *In Proc. NIST Performance Metrics for Intelligent Systems Workshop, Washington, DC, USA.*

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

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(6), 1296–1308.

Rybski, P. E., Yoon, K., Stolarz, J., & Veloso, M. M. (2007, March 10–12). Interactive robot task training through dialog and demonstration. *Proceeding of the ACM/IEEE 2nd International Conference on Human–robot Interaction,* (pp. 49–56) Arlington, Virginia, USA.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Sierhuis, M., Bradshaw, J. M., Acquisti, A., van Hoof, R., Jeffers, R., & Uszok, A. (2003). Human–agent teamwork and adjustable autonomy in practice. *Proceedings The 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space, Nara, Japan.*

Stiehl, W. D., Lieberman, J., Breazeal, C., Basel, L., Lalla, L., & Wolf, M. (2005). Design of a therapeutic robotic companion for relational, affective touch. *Proceedings 14th IEEE International Workshop on Robot and Human Interactive Communication*, (pp. 408–415). Nashville, USA.

Tajfel, H. (1982). *Social identity and intergroup behavior*. Cambridge, England: Cambridge University Press.

Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., et al. (2004). Is an entertainment robot useful in the care of elderly people with severe dementia? *Journals of Gerontology Series A: Biological and Medical Sciences, 59*(1), 83–85.

Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., & Schultz, A. C. (2005). Enabling effective human–robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man and Cybernetics* (Part A), *35*(4), 460–470.

Trafton, J. G., Schultz, A. C., Perznowski, D., Bugajska, M. D., Adams, W., Cassimatis, N. L., & Brock, D. P. (2006). Children and robots learning to play hide and seek. Paper presented at the ACM SIGCHI/SIGART Human-Robot Interaction, Salt Lake City, Utah, USA.Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction Studies, 8*(3). (This issue)

Turner, J. C. (1982). *Social identity and intergroup behavior*. Cambridge, England: Cambridge University Press.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (Vol. 2, pp. 77–121). Greenwich, CT: JAI Press.

Walsh, J. P., & Fahey, L. (1986). The role of negotiated belief structures in strategy making. *Journal of Management,* 12 (3), 325–338.

Walster, E., Walster, G. W., Berscheid, E., & Austin, W. (1978). *Equity: theory and research.* Boston: Allyn and Bacon.

Werry, I., Dautenhahn, K., Ogden, B., & Harwin, W. (2001). Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy. In M. Beynon, C. L. Nehaniv & K. Dautenhahn (Eds.), Proceedings of The Fourth International Conference on Cognitive Technology: Instruments of Mind (pp. 57-74). Berlin, Heidelberg: Springer-Verlag.

Wilson, E. O. (2000). *Sociobiology: the new synthesis* (revised ed.). Cambridge, MA: Harvard University Press.

Wilson, E. O., & Holldobler, B. (1990). *The ants*. Cambridge, MA: Harvard University Press.

Authors' address

Victoria Groom and Clifford Nass
Department of Communication
Stanford University
Stanford, CA 94305-2050

{vgroom,nass}@stanford.edu

About the authors

Victoria Groom is pursuing a Ph.D. in Communication from Stanford University. She received a B.A. in English and a M.A. in Media Studies from Stanford University in 2001. She is a member of the CHIMe Lab, where she researches issues of human–robot interaction and social responses to mediated communication.

Clifford Nass is the Thomas M. Storke Professor at Stanford University (2006–). He received a B.A. in mathematics and an M.A. and Ph.D. in sociology from Princeton University in 1981, 1985, and 1986, respectively. His research interests include communication between humans and interactive media, social responses to communication technology, and voice interfaces.