PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Role of Expressive Behavior for Robots that Learn from People

Journal:	*Philosophical Transactions B*
Manuscript ID:	RSTB-2009-0157
Article Type:	Research
Date Submitted by the Author:	10-Jul-2009
Complete List of Authors:	Breazeal, Cynthia; Massachusetts Institute of Technology, Media Lab
Issue Code: Click here to find the code for your issue.:	EMOTIONS
Subject:	Behaviour < BIOLOGY, Cognition < BIOLOGY
Keywords:	Social Robotics, Human robot interaction, Robot Learning, Expressive Behavior, Affective Computing

The Role of Expressive Behavior for Robots that Learn from People

Cynthia Breazeal

ABSTRACT

Robotics has traditionally focused on developing intelligent machines that can manipulate and interact with objects. The promise of personal robots, however, challenges researchers to develop socially intelligent robots that can collaborate *with* people to do things. In the future, robots are envisioned to assist people with a wide range of activities such as domestic chores, helping elders to live independently longer, serving a therapeutic role to help children with autism, assisting people undergoing physical rehabilitation, and much more. Many of these activities shall require robots to learn new tasks, skills, and individual preferences while “on the job” from people with little expertise in the underlying technology. This paper identifies four key challenges in developing social robots that can learn from natural interpersonal interaction. The author highlights the important role that expressive behavior plays in this process, drawing examples from the past eight years of her research group, the Personal Robots Group at the MIT Media Lab.

INTRODUCTION

Studies by the United Nations Economic Commission and International Federation of Robotics forecast a dramatic increase in consumer demand for robots that assist, protect, educate and entertain over the next twenty to thirty years. In the future, personal robots will be able to help people as capable assistants in their daily activities. Consider cooperative activities, such as preparing a meal together, building a structure with teammates, or teaching someone a new skill. Through sophisticated forms of social interaction and learning, people are able to accomplish more than they could alone. Socially intelligent robots could have significant positive impact on real-world challenges, such as helping elders to live independently at home longer, serving as learning companions for children and enriching learning experiences through play, serving a therapeutic role to help children with autism learn communication skills, or functioning as effective members of human-robot teams for disaster response missions, construction tasks, and more.

Many of these applications require robots to engage humans in sophisticated forms of social interaction, including human-centered multi-modal communication, teamwork, and social forms of learning such as tutelage. Over the past several years, my research has focused on endowing autonomous robots with social intelligence to enable them to engage in the powerful, social forms of interaction and learning that people readily participate. This vision is motivated by the observation that humans are ready-made experts in social interaction; the challenge is to design robots to participate in what comes naturally to people. By doing so, socially interactive robots could help not only specialists, but anyone.

Today, however, autonomous and semi-autonomous robots are widely regarded as tools that trained operators command and monitor to perform tasks. Beyond robustness and proficiency in the physical world, however, the promise of personal robots that can partake in the daily lives of people is pushing robotics and A.I. research in new directions. Whereas robotics has traditionally focused on developing machines that can manipulate and interact with *things*, the promise of personal robots challenges us to develop robots that are adept in their interactions with *people*. Further, in contrast to the traditional view of robots as sophisticated tools that we use to do things *for* us, this new generation of socially intelligent robots is envisioned as partners that collaborate to do things *with* us.

Over the past several years, new research fields have emerged (i.e., Human-Robot Interaction and Social Robotics) to address challenges in building robots that are skillful in their interactions with people [1-4]. Given that social robots are designed to interact with people in human-centric terms within human environments, many are humanoid (e.g., [5-6]), or animal-like (e.g., [7-8]) in form, and even the more mechanical-looking robots tend to have anthropomorphic movement or physical features (e.g., [9-10]).

A unifying characteristic is that social robots communicate and coordinate their behavior with humans through verbal, non-verbal, or affective modalities. For instance, these might include whole-body motion (e.g., dancing [11], walking hand-in-hand [12]), proxemics (i.e., how a robot should approach a person [13], follow a person [14], or maintain appropriate interpersonal distance [15]), gestures (e.g., pointing, shrugging shoulders, or shaking hands [16-18]), facial expressions (e.g., [19-22]), gaze behavior (e.g., [23-25]), head orientation and shared attention (e.g., [26-27]), linguistic and paralinguistic cues (e.g., [28-29]) or emotive vocalization (e.g., [30-31]), social touch-based communication (e.g., [32]), and how these cues complement verbal communication [e.g., 33].

Progress continues in building robots that can learn from people, either through observation, imitation or direct tutelage (for reviews see [34-35]). For instance, impressive strides have been made in designing robots that learn new skills (e.g., pendulum swing-up [36], body schema [37], peg insertion [38], dance gestures [39], communication skills and protocols [40-42]) as well as tasks (e.g., stacking objects [43-44], fetch and carry [45], setting a table [46], or sorting objects into bins [47,48]).

Modern robots are beginning to participate as members of heterogenous teams that cooperate with people to achieve shared goals. For instance, a remote human might supervise a distributed team of robots to perform a task (e.g., disaster response or search and rescue [49-50]). In addition, co-located teamwork has been explored such as a human and a robot working side-by-side [51], or a team of humans and robots working in the same area to assemble a structure [52].

Furthermore, as people begin to interact with robots more closely, it is important that robots' behavior, rationale, and motives be easily understood. The more these mirror natural human analogs, the more intuitive it becomes for us to communicate and coordinate our behavior with robots. Researchers have begun to explore the role of affect (e.g., [3, 53-55]), perspective taking and theory of other minds (e.g., [56-58]), and even simple forms of empathy [59-60] and models of attachment [61] in generating a robot's behavior.

A relevant issue underlying these different kinds of interactions is how people form social judgments of robots --- are robots perceived as trustworthy, persuasive, reliable, likeable, etc. (e.g., [62-63])? A number of groups have also explored how people's social judgments of robots compare to animated agents and even mixed-reality agents [64]. It is intriguing that the physical presence of robots seems to matter to people as robots often score higher than their virtual counterparts on measures of engagement, social presence, working alliance, as well as social influence on human behavior (e.g., [65-67]). Researchers have started delving into fMRI studies to try to understand these differences and to what extent people attribute human characteristics to robots, including theory of mind [68].

ROBOTS THAT LEARN FROM PEOPLE

Within this broader context of HRI and Social Robotics, this paper summarizes the past 8 years of research from my group (the Personal Robots Group at the MIT Media Lab [69-71]) with respect to significant lessons we have learned in our quest to build robots that can learn from anyone. My group is recognized for pioneering Human-Robot Interaction and Social Robotics through the development of expressive autonomous robots that socially interact with people in a natural manner [71]. Figure 1 presents the three "flagship" social robots we have developed starting with *Kismet* in the late 1990s, *Leonardo* spanning early-mid 2000s, and our new robot, *Nexi*. Each design is considered state-of-the-art (building upon lessons and technologies of its predecessor) and supports a different set of highly related scientific questions at the intersection of emotion and human-robot interaction, social learning, sophisticated forms of social cognition, and human-robot teamwork.

Figure 1: Three examples of social robots used in this research: Kismet (left), Leonardo (center), and Nexi (right).

One of my main research interests has been to develop robots that can learn from natural interpersonal interactions. Personal robots of the future will need to quickly learn new tasks and skills from people who are not specialists in robotics or machine learning techniques but possess a lifetime of experience in teaching and learning from one another. A major technical goal is to engineer robots that can leverage social guidance to efficiently and robustly acquire new capabilities from natural human instruction and to do so dramatically faster than it could alone. As an integral part of this endeavor, my group has contributed new knowledge and findings towards how humans teach social robots, and the important role that the robot's expressive behavior plays in this interpersonal process.

In contrast to traditional statistical machine learning approaches that require human expertise to craft a successful large-scale search problem that uses little or no real-time human input, my group's approach recognizes the advantage of designing robots that can leverage from the same rich forms of social interaction that people readily use to teach or learn from one another. Human teachers verbally and non-verbally guide the exploration of learners by directing attention, providing feedback, structuring experiences, supporting learning attempts, and regulating the complexity and difficulty of information to push learners a little beyond their current abilities in order to help them acquire new skills and concepts. In turn, learners tune their teachers' instruction and shape subsequent guidance by expressing their current understanding through demonstration and a rich variety of communicative acts. Through this interaction, learner and teacher from mental models of each other that they use to support the learning-teaching process as a richly *collaborative* activity.

It is actually quite difficult to build robotic systems that can successfully learn in real-time from the general public. Human teaching behavior is highly variable and complex, and different people bring different styles of interaction to the table. Today, it is common practice for robots to be taught and evaluated by the same researchers who developed it. Not surprisingly, if the teacher has special technical expertise and knowledge of the underlying learning algorithms that the robot uses, this leads to a strongly machine-centric style of interaction that is neither natural nor intuitive to someone who lacks such expertise. In fact, although there exists quite substantial work in developing robots that learn from people, it is still uncommon to conduct human participant studies with members of the general public to assess the learning performance of a robot when taught by someone who is *not* an expert in robotics, machine learning, or otherwise.

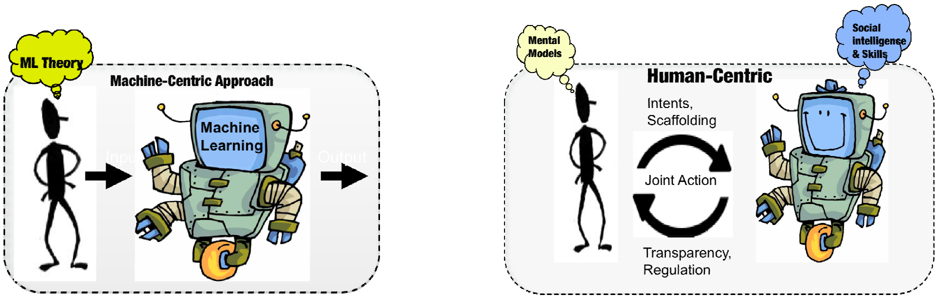


Figure 2: The traditional, machine-centric approach to teaching robots where the teacher is often has expertise in the robot's learning algorithms (left). Our new, human-centric approach that supports how "ordinary" people approach the task of teaching robots new skills or knowledge (right).

My research group is unusual for a robotics group, having conducted over a dozen controlled, in-lab human participant studies with hundreds of participants, to gain greater qualitative and quantitative understanding for how people approach the task of teaching a socially responsive machine. Often, we begin an investigation with a human study to learn more detail about how people teach each other. Then, computationally modeling this process allows us to identify and explore the use of a variety of social cues, expressive behaviors, skills and cognitive capabilities that support social learning in robots. In this way, we use social robots as a scientific tool for measuring and quantifying human behavior in new ways. This in turn has allowed us to generate new findings and discover new knowledge that can even inform how people teach and learn from one another. Figure 2 contrasts our human-centric model (right) to the traditional machine-centric view (left).

CHALLENGES IN BUILDING TEACHABLE ROBOTS

Applying these results, my group has developed and evaluated how these social behaviors and expressive capabilities enable robots to learn interactively with human participants, as well as how the same social skills address several key challenges in learning from natural human instruction. I highlight several challenges below together with research highlights of my research group's contributions toward their solution.

CHALLENGE 1: Robots face the situation that there is a fundamental mismatch in their social and communicative sophistication relative to humans. For effective learning, however, it is important that learners be slightly challenged to push themselves toward new abilities that are within reach, while avoiding situations where they are too overwhelmed to make sense of things. Fortunately, teachers and learners can work together to establish to a suitable level of difficulty and to regulate the complexity of the interaction to be suitable for both.

Example: Envelope Displays. To address this challenge, our research has contributed evidence for the importance of paralinguistic communication cues in human-robot interaction, and how they can be used to successfully manage this imbalance in a natural and intuitive manner. Through HRI studies with our robot, Kismet, we have found that humans readily entrain to a robot's non-verbal social cues (e.g., *envelope displays* that regulate the exchange of speaking turns in human conversation) to improve the efficiency and robustness of "conversational" flow by intuitively slowing the rate of turn exchanges to a level the robot can handle well. For instance, humans tend to make eye contact and raise their eyebrows when ready to relinquish their speaking turn, and tend to break gaze and blink when starting their speaking turn. When these same facial displays are implemented on a robot, we have found that they are effective in smoothing and synchronizing the exchange of speaking turns with human subjects, resulting in fewer interruptions and awkward long pauses between turns [72].

Example: Coordination Behaviors. Through another series of HRI studies, we have examined the use of a number of coordination behaviors where participants guided our robot, Leonardo, using speech and gesture to perform a physical task involving pressing a sequence of colored buttons ON. Leonardo communicates through gaze (visual attention) and facial expressions (affective state) or explicitly through gestural cues (i.e., pointing). The robot's coordination behaviors include visually attending to the human's actions (e.g., pointing to or pressing a button) to acknowledge their contributions, issuing a short nod to acknowledge the success and completion of the task or subtask (i.e., turning the buttons ON), visually attending to the person's attention directing cues such as to where the human looks or points, looking back to the human once the robot presses a button to make sure its contribution is acknowledged, and pointing to buttons in the workspace to direct the human's attention toward them. Both self-report via questionnaire and behavioral analysis of video support the hypothesis that these non-verbal communication cues positively impact human-robot task performance with respect to understandability of the robot, efficiency of task performance, and robustness to errors that arise from miscommunication [73].

Example: Emotive Displays. In addition, we have found that emotive expressions (as governed by the robot's emotion-based models) are interpreted by humans as natural analogs, and thereby can be used by the robot to regulate its interaction with the human---to keep the complexity of the interaction within the robot's perceptual limits, and even to help the robot to achieve its goals [74]. Many of these results were first observed with our robot, Kismet, the first robot designed to explicitly explore socio-emotive face-to-face interactions with people [71]. Our research with Kismet was strongly inspired by the origins of social interaction and communication in people, namely that which occurs between caregiver and infant, though extensive computational modeling guided by insights from developmental psychology and behavioral models from ethology [75]. It is well established that early infant-caregiver exchanges are grounded in the regulation of emotion and its expression.

Inspired by these interactions, Kismet's cognitive-affective architecture was designed to implement core proto-social responses exhibited by infants given their critical role in normal social development. Internally, Kismet's models of emotion interacted intimately with its cognitive systems to influence behavior and goal arbitration. Through a process of behavioral homeostasis, these emotive responses served to restore the robot's internal affective state to a mildly aroused, slightly positive state – corresponding to a state of interest and engagement in people and its surroundings that fosters learning. One purpose of Kismet's emotive responses was to reflect the degree to which its drives and goals were being successfully met. A second purpose was to use emotive communication signals to regulate and negotiate its interactions with people. Specifically, Kismet utilized emotive displays to regulate the intensity of playful interactions with people, making sure to keep the complexity of the perceptual stimulus within a range the robot could handle and potentially learn from. In effect, Kismet socially negotiated its interaction with people via its emotive responses to have humans help it achieve its goals, and satiate its drives, and maintain a suitable learning environment [76].

Summary: Joint Action. While more established approaches to instructing robots views the interaction as a one-way flow of information from human to machine, this body of work challenges this paradigm by illustrating the myriad of ways that humans participate in the teaching-learning process as *tightly coupled joint action*. Humans do not simply provide training inputs as a one-sided interaction to which the learner must react. Rather, people are constantly reading and interpreting numerous behavioral cues of the robot as indicators of its internal state, and are continually adapting and tuning their teaching behavior to be suitable for the robot learner.

This interaction dynamic has significant implications for the design of robots that learn from people. The robot is not restricted to learning in complex environment that does not care whether the robot succeeds or fails – a common assumption in robot learning systems. Rather, people view teaching and learning as a partnership with shared goals. Because of this, the robot can *proactively* improve the quality of its learning environment, tuning the teaching acts of the human to be more suitable, through using communication acts that reveal its learning process to the human teacher.

CHALLENGE 2: Faced with an incoming stream of sensory data, a robot must figure out which of its myriad of perceptions are relevant to the task at hand. This is an important capability for generating coherent behavior as well as for learning given that the search over state space becomes enormous as perceptual abilities and complexity of the environment increases.

Example: Saliency and Shared Attention. To address this challenge we have identified a set socially embodied cues and socio-cognitive abilities that assist the robot's determination of saliency when learning a task. These cues and abilities make the robot's underlying attention mechanisms responsive to a human teacher's efforts to highlight a distinct environmental context or change that is relevant to the learning task. In a series of human studies we have identified a growing set of social cues and socio-cognitive skills that play an effective role in addressing the saliency question.

For instance, we have implemented a multi-modal attention system to enable the robot to leverage the human teacher's desire to direct its visual attention by following the human's pointing gestures or gaze (estimated by head pose). To compute our robot's attentional focus, the attentional system computes the level of saliency (a measure of "interest") per feature channel for objects and events in the robot's perceivable space [77-78]. For Leonardo, the contributing factors to an object's overall saliency fall into three categories: its perceptual properties (i.e., its proximity to the robot, its color, whether it is moving, etc.), the internal state of the robot (i.e., whether this is a familiar object, what the robot is currently searching for, and other goals), and social reference (if something is pointed to, looked at, talked about, or is the referential focus). For each item in the perceivable space, the overall saliency at each time step is the result of the weighted sum for each of these factors. The item with the highest saliency becomes the current attentional focus of the robot, and also determines where the robot's gaze is directed. The gaze direction of the robot is an important communication device to the human, verifying for the human partner what the robot is attending to and thinking about.

The human's attentional focus is determined by what he or she is currently looking at. Leonardo calculates this using the head pose tracking data, assuming that the person's head orientation is a good estimate of his/her gaze direction. By following the person's gaze, the shared attention system determines which (if any) object is the attentional focus of the human's gaze. The mechanism by which infants track the referential focus of communication is still an open question, but a number of sources indicate that looking time is a key factor, such as word learning studies. For example, when a child is playing with one object and hears an adult say "It's a modi", the child does not attach the label to the object the child happens to be looking at (which is often the adults face!). Instead the child redirects his or her attention to look at what the adult is looking at, and attaches the label to that object. For our robot, we use a simple voting mechanism to track a *relative-looking-time* for each of the objects in the robot's and human's shared environment. The object with the highest accumulated *relative-looking-time* is identified as the referent of the communication between the human and the robot [79].

Using these models, we have found that active monitoring of shared visual attention between human teacher and robot learner is important to achieve robustness in the learning interaction. In a series of human participant studies where human teachers guide a robot to perform a simple task (learning to operate a control panel with a lever, toggle, and button), we have found that humans readily coordinate their teaching behavior with the robot's gaze behavior --- waiting until the robot re-establishes eye-contact before offering their next guidance cue, adaptively re-orienting their guidance cue to be in alignment with the robot's current visual focus, actively trying to re-direct the robot's gaze though deictic cues, or offering more guidance if the robot's gaze behavior conveys uncertainty in what to do next (e.g., looking back and forth among several possible alternatives) [80-81]. These findings suggest that people read the robot's gaze as an indicator of its internal state of attention as well as solicitations for help, and intuitively coordinate their teaching acts to support the robot's learning process.

Example: Perspective Taking. In another series of human and HRI studies, we identified, verified, and evaluated mental perspective taking as an important socio-cognitive skill that helps either human or robot learners to focus attention on the subset of the problem space that is important to the teacher by actively considering the teacher's experience such as visual perspective, attentional focus, or resource considerations [82]. This constrained attention enables the robot learner to overcome ambiguity and incompleteness that is often present in human demonstrations.

To endow Leonardo with perspective taking abilities, our cognitive-affective architecture incorporates simulation-theoretic mechanisms as a foundational and organizational principle. *Simulation Theory* holds that certain parts of the brain have dual use; they are used to not only generate our own behavior and mental states, but also to predict and infer the same in others. To try to recognize or infer another person's mental process, the robot uses its own cognitive processes and body structure to simulate the mental states of the other person -- in effect, taking the mental perspective of another.

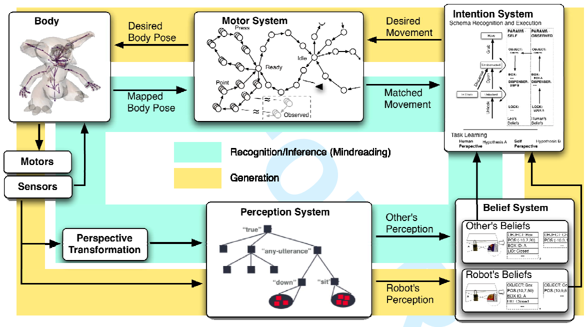


Figure 3: The Self-As-Simulator Architecture.

In Figure 3, the two concentric bands denote two different modes of operation. In *generation mode* (the light band) the robot constructs its own mental states to behave intelligently in the world. In *simulation mode* (the dark band) the robot constructs and represents the mental states of its human collaborator based on observing his or her behavior and taking their mental perspective. By doing so, the mental states of the human and robot are represented in the same terms so they can be readily compared and related to one another. For instance, within the Perception System, the robot performs a transformation to estimate what the human partner can see from his or her vantage point. Within the Motor System, mirror-neuron inspired mechanisms are used to map and represent perceived body positions of the human into the robot's own joint space to conduct action recognition. Within the Belief System, belief-construction is used in conjunction with adopting the visual perspective of the human partner in order to estimate the beliefs the human is likely to hold given what he or she can visually observe. Finally, within the Intention System where goal-directed behaviors are generated, schemas relate preconditions and actions with desired outcomes and are organized to represent hierarchical tasks. Within this system, motor information is used along with perceptual and other contextual clues (i.e., task knowledge) to infer the human's goals and how he or she might be trying to achieve them (i.e., plan recognition).

In a learning situation, the robot can take the perspective of the teacher in order to model the task from their perspective. In effect, the robot runs a parallel copy of its task-learning engine that operates on its simulated representation of the human's beliefs. In essence, this focuses the hypothesis generation mechanism on the subset of the input space that matters to the human teacher. This enables the robot to learn what the teacher intends to teach even if the demonstrations are ambiguous.

To investigate this, we conducted a human participant study where the participants were asked to engage in four different learning tasks involving foam building blocks. We gathered data from 41 participants, divided into two groups. 20 participants observed demonstrations provided by a human teacher sitting opposite them (the social condition), while 21 participants were shown static images of the same demonstrations, with the teacher absent from the scene (the nonsocial condition). Participants were asked to show their understanding of the presented skill either by re-performing the skill on a novel set of blocks (in the social context) or by selecting the best matching image from a set of possible images (in the nonsocial context). Figure 4 (left) illustrates sample demonstrations of each of the four tasks. The tasks were designed to be highly ambiguous, providing the opportunity to investigate how different types of perspective taking might be used to resolve these ambiguities. The subjects' demonstrated rules can be divided into three categories: perspective taking (PT) rules, non-perspective taking (NPT) rules, and rules that did not clearly support either hypothesis (Other). For instance, *Task 1* focused on visual perspective taking during the demonstration. Participants were shown two demonstrations with blocks in different configurations. In both demonstrations, the teacher attempted to fill all of the holes in the square blocks with the available pegs. Critically, in both demonstrations, a blue block lay within clear view of the participant but was occluded from the view of the teacher by a barrier. The hole of this blue block was never filled by the teacher. Thus, an appropriate (NPT) rule might be "fill all but blue," or "fill all but this one," but if the teacher's perspective is taken into account, a more parsimonious (PT) rule might be "fill all of the holes" (see Figure 4).

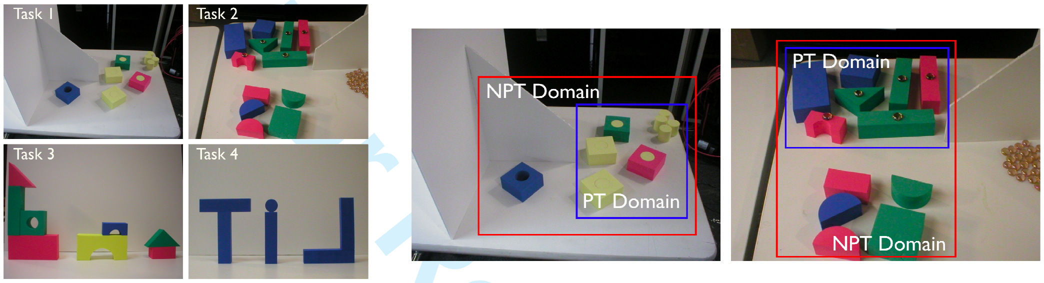
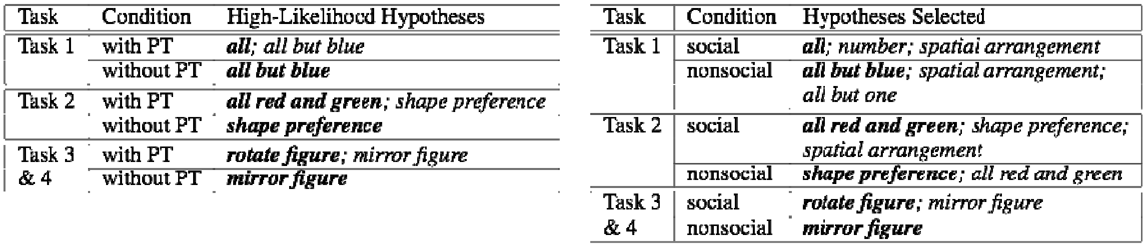


Figure 4: The four tasks presented to human participants. In task 1, subjects were asked to infer the rule for which blocks received a yellow peg. A visual occlusion presents a different viewpoint of the demonstration between teacher and learner (the teacher can not see the hidden blue block). In task 2, subjects were asked to infer the rule for which blocks get a bead where proximity and gaze were used to denote which blocks the teacher intended to use for the demonstration. Task 3 and 4 were figure assembly tasks. (Right) shows different sets of blocks that that could be considered when learning the rule if perspective taking (PT) is used, or not (NPT).

The tasks from our human study were used to create a benchmark suite for our architecture. In our simulation environment, the robot was presented with the same task demonstrations as were provided to the study participants. The learning performance of the robot was analyzed in two conditions: with the perspective taking mechanisms intact, and with them disabled. Table 1 (left) shows the hypotheses entertained by the robot in the various task conditions at the conclusion of the demonstrations. The hypotheses favored by the learning mechanism are highlighted in bold. For comparison, Table 1 (right) displays the rules selected by study participants, with the most popular rules for each task highlighted in bold. For every task and condition, the rule learned by the robot matches the most popular rule selected by the humans.



Task	Condition	High-Likelihood Hypotheses
Task 1	with PT	***all***; *all but blue*
	without PT	***all but blue***
Task 2	with PT	***all red and green***; *shape preference*
	without PT	***shape preference***
Task 3 & 4	with PT	***rotate figure***; *mirror figure*
	without PT	***mirror figure***

Task	Condition	Hypotheses Selected
Task 1	social	***all***; *number; spatial arrangement*
	nonsocial	***all but blue***; *spatial arrangement; all but one*
Task 2	social	***all red and green***; *shape preference; spatial arrangement*
	nonsocial	***shape preference***; *all red and green*
Task 3 & 4	social	***rotate figure***; *mirror figure*
	nonsocial	***mirror figure***

Table 1: (Left) rules learned by the robot with perspective taking enabled (PT) or disabled. (Right) the corresponding rules learned by people for the same tasks. The difference in rule choice by subjects between social and nonsocial condition is highly significant ($p < 0.001$).

These results support our hypothesis that the robot's perspective taking mechanisms focus its attention on a region of the input space similar to that attended to by study participants in the presence of a human teacher. It should also be noted, as evident in Table 1, participants generally seemed to entertain a more varied set of hypotheses than the robot. In particular, participants often demonstrated rules based on spatial or numeric relationships between the objects --- relationships that are currently not yet represented by the robot. Thus, the differences in behavior between the humans and the robot can largely be understood as a difference in the scope of the relationships considered between the objects in the example space, rather than as a difference in this underlying space. The robot's perspective taking mechanisms seem to be successful at bringing the robot's focus of attention into alignment with the humans' in the presence of a social teacher.

Example: Spatial Scaffolding: In other human participant and HRI experiments, we have identified, verified, and evaluated a set of simple, prevalent, and highly reliable spatial scaffolding cues by which human teachers interactively structure and organize the physical workspace to help direct the attention of the learner (e.g., moving objects nearer or farther from the learner's body to signify their relevance) [83].

For example, we designed set of tasks to examine how teachers emphasize and de-emphasize objects in a learning environment with their bodies, and how this emphasis and de-emphasis guides the exploration of a learner and ultimately the learning that occurs. In our human study, we gathered data from 72 individual participants, combined into 36 pairs. For each pair, one participant was randomly assigned to play the role of teacher and the other participant assigned the role of learner for the duration of the study. For all of the tasks, participants were asked not to talk, but were told that they could communicate in any way that they wanted other than speech. The teacher and learner stood on opposite sides of a tall table, with 24 colorful foam building blocks (four different colors and six different shapes) arranged between them on the tabletop. The study tasks were interactive "secret constraint" tasks where one person (the learner) knows the task goal (construct a tangram-like figure out of the blocks) but does not know that there is secret constraint to accomplish the task successfully. The other person (the teacher) does not know the task goal (the figure) but knows the constraint (e.g., "the figure must be constructed using only blue and red blocks, and no other blocks."). Hence, both people must work together to successfully complete the task.

To record high-resolution data of the study interactions, we developed a data-gathering system which incorporated multiple, synchronized streams of information about the study participants and their environment. For all of the tasks, we tracked the positions and orientations of the heads and hands of both participants, recorded video of both participants, and tracked all of the objects with which the participants interacted such as the positions and orientations of all of the foam blocks. To identify the emphasis and de-emphasis cues provided by the teachers in these tasks, an important piece of "ground-truth" information was exploited: for these tasks, some of the blocks were "good," and others of the blocks were "bad." In order to successfully complete the task, the teacher needed to encourage the learner to use some of the blocks in the construction of the figure, and to steer clear of some of the other blocks.

We observed a wide range of embodied cues provided by the teachers in the interactions for these two tasks, as well as a range of different teaching styles. Positive emphasis cues included simple hand gestures such as tapping, touching, and pointing at blocks with the index finger. These cues were often accompanied by gaze targeting, or looking back and forth between the learner and the target blocks. Other positive gestures included head nodding, the "thumbs up" gesture, and even shrugging. Teachers nodded in accompaniment to their own pointing gestures, and also in response to actions taken by the learners. Negative cues included covering up blocks, holding blocks in place, or maintaining prolonged contact despite the proximity of the learner's hands. Teachers would occasionally interrupt reaching motions directly by blocking the trajectory of the motion or even by touching or (rarely) lightly slapping the learner's hand. Other negative gestures included head shaking, finger or hand wagging, or the "thumbs down" gesture.

However, by far the most important set of cues used related to block movement and the use of space. To positively emphasize blocks, teachers would move them towards the learner's body or hands, towards the center of the table, or align them along the edge of the table closest to the learner. Conversely, to negatively emphasize blocks, teachers would move them away from the learner, away from the center of the table, or line them up along the edge of the table closest to themselves. Teachers often devoted significant attention to clustering the blocks on the table, spatially grouping the bad blocks with other bad blocks and the good blocks with other good blocks. These spatial scaffolding cues were the most prevalent cues in the observed interactions [83].

To verify the prevalence and usefulness of these spatial scaffolding cues for a robot, we substituted our robot Leonardo for the role of the learner [84]. The robot's attention system was designed to pay attention to block movement toward and away from its body. In order to give the robot the ability to learn from these embodied cues, we developed a simple, Bayesian learning algorithm. The algorithm was designed to learn rules pertaining to the color and shape of the foam blocks and maintained a set of classification functions which tracked the relative odds that the various block attributes were "good" or "bad" according to the teacher's secret constraints. Each time the robot observed a salient teaching cue, these classification functions were updated using the posterior probabilities presented in the previous section - the odds of the target block being "good" or "bad" given the observed cue. For example, if the teacher moved a green triangle away from the robot, the relative odds of *green* and *triangular* being good block attributes would decrease. Similarly, if the teacher then moved a red triangle towards the robot, the odds of *red* and *triangular* being good would increase.

These simple spatial scaffolding cues proved to be highly effective. We invited 18 participants to teach Leonardo the identical secret constraint tasks as our human learners. The robot successfully learned the task in 33 of the 36 interactions (92%). These results support the conclusion that the spatial scaffolding cues observed in human-human teaching interactions do indeed transfer to human-robot interactions, an can be effectively taken advantage of by robot learners [84].

Summary: Social Filters. Whereas traditional approaches to teaching robots do not model social-cognitive skills and abilities as integral to the learning process, this body of work has identified and verified a number of ways that internal and external social factors play an important role in how a robot learner filters the incoming perceptual stream to attend to what matters, that human teachers bring many of these same social cues and skills to bear when teaching either humans or robots, and that these "*social filters*" can be effectively used by a robot to help it identify the most relevant items to consider, thereby making the learning problem significantly more manageable.

CHALLENGE 3: Once the robot has identified salient aspects of the scene, how does it determine what actions it should take? If the robot had a way of focusing on potentially successful actions, its exploration would be more effective. This can be addressed in a number of ways, such as by experimenting on its own *as in reinforcement learning. However, for large state-action spaces this typically requires a prohibitively large number of trials.*

Example: Tutelage-Style Interaction: To address this issue, we have explored how social skills such as turn-taking enables a human teacher to play an important and flexible role in guiding the robot's exploration. This focuses the robot's selection of the most promising actions in specific contexts to discover solutions more quickly. By participating in a "dialog" of demonstration followed by feedback and refinement, the human helps the robot to determine what action to try through a communicative and iterative process. We evaluated this approach by comparing it to learning the same task using traditional reinforcement learning and achieved significant improvements in efficiency without loss of accuracy and with decreased sensitivity to noise (e.g., errors introduced by miscommunication are quickly repaired which leads to greater robustness) [85].

Example: Socially Guided Exploration: Unfortunately, a common limitation of human teachable robots is that the robot only learns when being explicitly taught. Personal robots, however, will need to learn while "on the job" even when a person is not present or willing to teach it. To address this, we have developed and evaluated a learning system whereby learning opportunities for the robot's hierarchical reinforcement learning mechanism arise from a combination of intrinsically motivated self-exploration and social scaffolding provided by a human teacher such as suggesting actions for the robot to try, drawing the robot's attention to relevant contexts, and highlighting interesting outcomes [86]. We have systematically identified and verified our set of social scaffolding mechanisms through a series of HRI studies where a human teacher guides Leonardo's exploration as it discovers a set of behaviors (e.g., opening or closing, playing music, changing colors of lights) of a "smart box" through pressing buttons, pushing levers, and sliding toggles. Over time, Leonardo learns a set of task policies for bringing about each of these behaviors from different starting conditions to "master" the "smart box". We analyzed the learning performance of the robot both with and without human teachers and found that learning performance via self-exploration is slower but more serendipitous resulting in a broader task suite, whereas learning with a human teacher makes learning more efficient and robust, but tends to result in a smaller, more specialized task suite that reflects what the person wanted the robot to learn [86].

Summary: Intrinsically Motivated but Guidable Learning: Personal robots will need to adapt their learning style to suit the dynamics of a changing learning environment. Sometimes the robot will have to explore on its own, while other times a teacher might be present to help guide the robot's exploration. Through our studies, we have found that each style of learning has its respective advantages and produces learning products that are synergistic. For instance, what is learned more slowly but serendipitously through intrinsically motivated exploration yields a broader task suite that can come in handy at later date -- - especially when the robot encounters a human teacher who helps the robot to rapidly hone and build on its growing skill set through socially guided exploration. Importantly, the mechanisms by which the robot's learning can be guided by the human should be informed by how people are naturally inclined to teach robots.

CHALLENGE 4: Once the robot attempts to perform an action, how can it determine whether it has been successful? How does it assign credit for that success? Further, if the robot has been unsuccessful, how does it determine which parts of its performance were inadequate? It is important that the robot be able to diagnose its errors in order to improve performance.

Example: Multi-Modal Feedback. To address this challenge, our approach recognizes that the teacher can readily help the robot do this given that he or she has a good understanding of the task and knows how to evaluate the robot's success and progress. One way that a human facilitates a learner's evaluation process is by providing feedback through various communication channels. For instance, we demonstrated the capability for a robot to interpret and appropriately respond to the affective intent in human speech, such as praising or scolding tones of voice [87]. In HRI studies we showed that people refer to the robot's expressive cues to confirm that the robot understood them as well as the strength of the affective intent. We have applied verbal feedback in teaching scenarios to help the robot correct its task model as soon as mistakes are made. Furthermore, the robot provides the human with communicative feedback so that misunderstandings can be detected quickly. Both forms of feedback help to prevent errors from persisting for multiple steps, which could make them more awkward to correct later on. In recent HRI studies, our data suggests that these various forms of feedback contribute to a more fluid, efficient, accurate, and robust teaching/learning interaction [85-86].

Example: Guidance and Understanding Intent. Note that for any given feedback channel, it is important to understand what people are trying to communicate through it and how they are trying to make use of it. Our HRI studies with an interactive reinforcement learning (RL) agent revealed that people use the reward signal to not only provide feedback on past actions (what is commonly assumed in the design of RL algorithms), but we discovered that people *also* use it to guide future action [81]. Further, we discovered a strong bias of positive over negative feedback over the entire duration of the training, even in the beginning when the agent was doing many things wrong [81]. This suggests that people were using the feedback channel to motivate and encourage the robot. In short, people were naturally inclined to use the reward signal in many ways that the traditional RL framework was not designed to handle. Given our findings, we were then able to adapt the RL agent algorithm and teaching interface to accommodate how and what people were trying to communicate to the learner. As a result, our modified RL agent learned much more efficiently and robustly in a subsequent series of HRI experiments [81].

Summary: Transparency. While traditional approaches to robot training do not consider how a robot can proactively communicate and reveal its learning process to the human teacher, the findings generated by this body of work argues for the importance of *transparency* in designing interactive robot learners. People are willing and able to help robots address the difficult task of assigning value to its past actions. People are also willing to help guide the robot to select good future actions, to motivate the robot, and more. However, human teachers cannot do this well if they lack a good mental model of the robot's learning process, or if not provided with the right set of communication channels. The robot's behavior, both its expressive cues and instrumental actions, can play a significant role in shaping the mental model that the human has for the robot. These readily observable expressive and performance-based cues make the robot's learning process transparent to the teacher. Much of our work to date has emphasized the role of the robot's non-verbal cues, such as facial expressions, gestures, and use of gaze, in supporting this process. And conversely, our HRI studies have helped us to identify what kinds of intents people want to communicate to the robot through both verbal and non-verbal channels --- to help the robot learn by influencing its evaluation process and more.

CONCLUSION

While it might be tempting to compare our outcomes with those of statistical machine learning techniques, my research vision and the challenges I wish to solve are ultimately different. My students and I have built and evaluated autonomous robotic systems that are able to leverage from the interplay of social guidance with statistical inference algorithms to learn new tasks and concepts from humans from natural social interactions. For task learning, our robots are able to quickly infer the critical preconditions and desired outcome for each step of the learned task, as well as how these steps relate to one another in the overall task structure, with improved efficiency and robustness to noise without loss of accuracy over traditional statistical machine learning methods (e.g., traditional reinforcement learning). For concept learning, our robots are able to learn the correct concept from natural interactions by exploiting natural scaffolding cues such as how the teacher uses space to highlight the concept to be learned, or by applying socio-cognitive skills to consider the teachers' perspective in order to learn the appropriate concept in the face of ambiguous demonstrations. The underlying machine learning algorithm can be simple because the robot appropriately leverages the social structure inherent in the teacher's behavior or the modified workspace to attend to what matters and learn the right thing. Furthermore, the same social cues can be repurposed to support other social capabilities such as multi-modal communication and our research on human-robot teamwork.

To conclude, the field of social robotics is very young but growing rapidly --- motivated by the vision of personal robots that help anyone in their daily activities. My dream is to enable machines to engage in the powerful, social forms of interaction, collaboration, understanding, and learning that people readily participate. This vision is motivated by the observation that humans are ready-made experts in social interaction; the challenge is to design robots to participate in what comes naturally to people. By doing so, socially interactive robots could help a wide demographic of people in a broad range of applications and real-world challenges from health, therapy, education, communication, security, entertainment, or physical assistance. In this article, I've tried to illustrate the myriad of ways in which designing social robots that successfully interact with and learns from ordinary people presents new challenges and opportunities, and highlighted some of the key lessons and findings I've learned along the way. We live in an exciting time where so much is possible at the intersection of science and technology. Social robots promise not only to be helpful to us in the future, but also a lot of fun. And in the process of building them, we may learn even more about ourselves.

REFERENCES

[1] Fong, T., Nourbakshsh,, I., Dautenhahn, K. (2003). "A Survey of Social Robots," *Robotics and Autonomous Systems,* **42**, pp. 143 – 166. 2003.

[2] Breazeal, C. (2004). "Social interactions in HRI: The robot view," *IEEE Transactions in Systems, Man, and Cybernetics*, *Part C*, 34(2), 181-186.

[3] Duffy, B.R. (2008). "Fundamental Issues in Affective Intelligent Social Machines," *The Open Artificial Intelligence Journal*, pp.21-34 (14), ISSN: 1874-0618 Volume 2, 2008

[4] Dautenhahn, K., "Getting to know each other – Artificial social intelligence for autonomous robots," *Robotics and Autonomous Systems*, vol. 16, pp. 333-356, 1995.

[5] Ogura, Y., Aikawa, H., Shimomura, K., Kondo, H., Morishima, A., Lim, H., Takanishi, A., "Development of a new humanoid robot WABIAN-2," in *Proc. IEEE International Conference on Robotics and Automation*, pp. 76-81, 2006.

[6] Tanaka, F., Noda, K., Sawada, T., Fujita, M., "Associated Emotion and its Expression in an Entertainment Robot QRIO", in *Proc. of the 3rd International Conference on Entertainment Computing*, (Eindhoven), pp. 499-504, 2004.

[7] Fujita, M., "On activating human communications with pet-type robot AIBO", in *Proc. of IEEE*, Vol. 92 (11), pp. 1804-1813, 2004.

[8] Wada, K., Shibata, T., Sakamoto, K., Tanie, K., "Long-term Interaction between Seal Robots and Elderly People — Robot Assisted Activity at a Health Service Facility for the Aged," in *Proc. of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment*, pp 325-330, 2005.

[9] Kozima, H., "An anthropologist in the children's world: A field study of children's everyday interaction with an interactive robot," in Proc. of the *International Conference on Development and Learning* (ICDL-2006; Bloomington, IN, USA), 2006.

[10] Tanaka, F., Movellan, J.R., Fortenberry B., and Aisaka K., "Daily HRI Evaluation at a Classroom Environment: Reports from Dance Interaction Experiments," in *Proc. of the 1st Annual Conference on Human-Robot Interaction (HRI 2006),* Salt Lake City, U.S.A., p.3-9, March 2006.

[11] Duffy, B. (2003). "Anthropomorphism and the Social Robot." *Robotics and Autonomous Systems,* Volume 42, Issues 3-4, Pages 177-190.

[12] Lim, H., Ishii, A., Takanishi, A., "Emotion-based Biped Walking", *Int. Journal of Information, Education and Research in Robotics and Artificial Intelligence*, Vol.22, No.5, pp. 577-586, 2004.

[13] Walters, M.L., Dautenhahn, K., Woods, S.N., Koay, K.L., “Robotic Etiquette: Results from User Studies Involving a Fetch and Carry Task,” in *Proc. of the* 2^{nd} *ACM/IEEE International Conference on Human-Robot Interaction*.

[14] Gockley, R., Forlizzi, J., Simmons, R., “Natural person-following behavior for social robots,” in *Proc. of Human-Robot Interaction*, March, 2007, pp. 17-24.

[15] Brooks, A. G., Arkin, R. C. (2007). “Behavioral overlays for non-verbal communication expression on a humanoid robot,” *Autonmous Robots*, vol. 22(1), pp. 55-74.

[16] Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M.C, Dario, P., Takanishi A., “Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII”, in *Proc. of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.2203-2208, 2004.

[17] Roccella, S., Carrozza, M.C., Cappiello, G., Dario, P., Cabibihan, J., Zecca, M., Miwa, H., Itoh, K., Matsumoto, M., Takanishi, A. (2004). “Design, fabrication and preliminary results of a novel anthropomorphic hand for humanoid robotics: RCH-1,” in *Proc. of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Sendai), pp.266-271.

[18] Miwa, H., Itoh, K., Takanobu, H., Takanishi, A. (2004). “Design and Control of 9-DOFs Emotion Expression Humanoid Arm”, in *IEEE International Conference on Robotics and Automation*, (New Orleans), pp.128-133, 2004.

[19] Iida, F., Tabata, M., Hara, F. (1998). “Generating personality character in a Face Robot through interaction with human, *Proc. of 7th IEEE International Workshop on Robot and Human Communication*, pp. 481-486,

[20] Berns, K., J. Hirth, J., “Control of facial expressions of the humanoid robot head ROMAN. *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Bejing, China, pp 3119-3124, 2006.

[21] Hayashi, K., Onishi, Y., Itoh, K., Miwa, H., Takanishi, A., “Development and evaluation of face robot to express various face shape,” in *Proc. of IEEE International Conference on Robotics and Automation*, pp. 481-486, 2006.

[22] DiSalvo, C., Gemperle, F., Forlizzi, J., Kiesler, S. (2002). “All robots are not created equal: the design and perception of humanoid robot heads”. Designing Interactive Systems. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques* Pages: 321 – 326.

[23] Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C., “Explorations in Engagement for Humans and Robots,” *Artificial Intelligence*, vol. 166(1-2), pp. 140-164, 2005.

[24] Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., Ikeuchi, K., “Flexible Cooperation between Human and Robot by interpreting Human intention from Gaze Information,” in *Proc. of the International Conference on Intelligent Robot and Systems*, pp. 846-851, 2004.

[25] Kikuchi, H., Yokoyama, M., Hoashi, K., Hidaki, Y., Kobayashi, T., Shirai, K., “Controlling Gaze of Humanoid in Communication with Human,” in *Proc. IROS*, pp.255-260, Oct. 1998.

[26] Fujie, S., Ejiri, Y., Nakajima, K., Matsusaka, Y., Kobayashi, T., “A conversation robot using head gesture recognition as paralinguistic information,” in *Proc. IEEE RO-MAN* 2004, pp.158-164, Sep. 2004.

[27] Imai, M., Ono, T., Ishiguro, H., “Physical relation and expression: Joint attention for human- robot interaction,” in *Proc. RO-MAN 2001*, pp. 512 – 517, 2001.

[28] Fujie, S., Fukushima, K., Kobayashi, T., “Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system,” in *Proc. Interspeech 2005*, pp.889-892, Sep. 2005.

[29] Matsusaka, Y., Tojo, T., Kobayashi, T., “Conversation robot participating in group conversation,” in *Trans. IEICE*, Vol. E86-D, No.1, pp.26-36, Jan. 2003.

[30] Cahn, J.E. (1990). "The generation of affect in synthesized speech", *J. American Voice Input/Output Society*, 8, 1-19.

[31] Abadjieva, E., Murray, I.R. and Arnott, J.L. (1993). "Applying analysis of human emotional speech to enhance synthetic speech", *Proc. Eurospeech ’93*, Berlin, Germany, pp. 909-912.

[32] Stiehl, W., Lieberman, J., Breazeal, C., Basel, L., Lalla, L., Wolf, M., “Design of a Therapeutic Robotic Companion for Relational, Affective Touch,” in *Proc. of 14th IEEE Workshop on Robot and Human Interactive Communication*, pp. 408-415, 2005.

[33] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (Eds.) *Embodied Conversational Agents*. Boston: MIT Press. 2000.

[34] Schaal, S. (1999) Is imitation learning the route to humanoid robots? *Trends in Cognitive Science* 3(6) pp. 233-242.

[35] Argall, B., Chernova, S. and Veloso, M.(2009). A Survey of Robot Learning from Demonstration. *Robotics and Autonomous Systems*. Vol. 57, No. 5, pages 469-483.

[36] Atkeson, C.G. and Schaal, S. (1997b) “Robot learning from demonstration,” *International Conference on Machine Learning*, pp. 11—73.

[37] Hersch, M., Sauser, E., and Billard, A. (2008). “Online learning of the body schema,”*International Journal of Humanoid Robotics,* 5 (2008) 161–181.

[38] Hovland, G.E. *et al.* (1996) “Skill acquisition from human demonstration using a hidden Markov Model,” in *IEEE International Conference on Robotics and Automation*, Minneapolis, MN, pp.2706—2711, IEEE.

[39] Maja J. Mataric, Victor B. Zordan, Zachary Mason (1998). “Movement Control Methods for Complex, Dynamically Simulated Agents: Adonis dance the Macarena”. *Proceedings Agents 1998*: 317-324

[40] Scassellati, B. (1998) **“**Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot,” in *Computation for Metaphors, Analogy and Agents* (C. Nehaniv, ed.), Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*, Springer-Verlag.

[41] Billard A. *et al.* (1998) “Experiments on human-robot communication with Robota, an imitative learning and communicating doll robot,” in *Proceedings of Socially Situated Intelligence Workshop as part of the Fifth Conference of the Simulation of Adaptive Behavior. Centre for Policy Modelling technical report series number CPM-98-38*

[42] Deb Roy and Alex Pentland. (1998). “Learning Audio-Visually Grounded Words from Natural Input,” *AAAI workshop on The Grounding of Word Meaning: Data and Models*, Madison, WI.

[43] Kuniyoshi, Y. *et al.* (1994). “Learning by watching: extracting reuseable task knowledge from visual observation of human performance*,” IEEE Trans. Robotics Automation* 10, 799—822.

[44] Calinon, S. and Billard, A. (2007). “What is the Teacher's Role in Robot Programming by Demonstration? - Toward Benchmarks for Improved Learning,” *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction*, 8.

[45] Monica N. Nicolescu, Maja J. Mataric (2003). “Natural methods for robot task learning: instructive demonstrations, generalization and practice,” AAMAS 2003: 241-248

[46] Pardowitz, M. Zoeliner, R., Knoop, S. & Dilmann, R. (2006). “Incremental learning of tasks from user demonstrations, past experiences and vocal comments,” *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, 36(5).

[47] Saunders, J., Nehaniv, C., Dautenhahn, K. (2006). “*Teaching robots by moulding behavior and scaffolding the environment*,” In *proceedings of the ACM SIGCHI/SIGART conference on human-robot interaction (HRI).* 118-125.

[48] Sonia Chernova and Manuela Veloso. (2008). “Teaching Collaborative Multi-Robot Tasks through Demonstration,” *In IEEE-RAS International Conference on Humanoid Robots*, Daejeon, Korea, December 2008.

[49] Bluethmann, W., R, Ambrose, M. Diftler, E. Huber, A, Fagg, M. Rosenstein, R. Platt, R. Grupen, C. Breazeal, A. Brooks, A. Lockerd, R. A.Peters II, O. C. Jenkins, M. Mataric, M. Bugajska (2004) “Building an Autonomous Humanoid Tool User.” *Proceedings of the IEEE/RAS Fourth International Conference on Humanoid Robots (Humanoids 2004)*. Santa Monica, CA. 402-421.

[50] Murphy, R.R., Tadokoro, S., and Fiorini, P. (2008). "Rescue Robotics," *Handbook of Robotics, A. Zelinkski, B. Sciliano*, editors, IEEE Press.

[51] J. A. Adams, C. M. Humphrey, M. A. Goodrich, J. L. Cooper, B. S. Morse, C. Engh and N. Rasmussen. Cognitive Task Analysis for Developing UAV Wilderness Search Support. *Journal of Cognitive Engineering and Decision Making*, 3(1), pp 1-26, 2009.

[52] Fong, T., Nourbakhsh, I., Ambrose, R., Simmons, R., Schultz, A., Scholtz, J. (2005). The Peer-to-Peer Human-Robot Interaction Project. In proceedings of *AIAA Space 2005*, September, 2005.

[53] Picard, R., (2000). *Affective Computing*. MIT Press.

[54] Fellous, J.-M. and Arbib, M. A. (eds.) (2005). *Who needs emotions?: the brain meets the robot.* Oxford University Press.

[55] Cañamero, L., Animating Affective Robots for Social Interaction. In L. Cañamero, R. Aylett (Eds.), Animating Expressive Characters for Social Interaction. John Benjamins Publishing Co., *Advances in Consciousness Research Series*.

[56] Scassellati, B. (2001). Theory of mind for a humanoid robot. *Autonomous Robots*, 12, 1. 13-24.

[57] 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 in Systems, Man and Cybernetics, Part A: Systems and Humans*, 35, 4. 460-470.

[58] Johnson, M., Demiris, Y., “Perceptual Perspective Taking and Action Recognition.” International Journal of Advanced Robotic Systems, 2(4), pp. 301-308, December 2005.

[59] Dautenhahn, K., “I could be you: The phenomenological dimension of social understanding,” *Cybernetics and Systems*, vol. 28, pp. 417-453, 1997.

[60] Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., Blumberg, B. (2005). "Learning from and about Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots," L. Rocha and F. Almedia e Costa (eds.), *Artificial Life*, 11(1-2), 1-32.

[61] Cañamero, L., Blanchard, A., Nadel, J. (2006). Attachment Bonds for Human-Like Robots. *International Journal of Humanoid Robotics,* 3(3): in press. ISSN: 0219-8436.

[62] Kidd, C., and Breazeal, C. (2008). "Robots at Home: Understanding Long-Term Human-Robot Interaction". Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008). Nice, France.

[63] Siegel, M. (2008). *Persuasive Robotics: Towards Understanding the Influence of a Mobile Humanoid Robot over Human Belief and Behavior*. September 2008. S. M. Media Arts and Sciences.

[64] Holz, T.; Dragone, M.; O'Hare, G.M.P. (2009). "Where Robots and Virtual Agents Meet: A Survey of Social Interaction Across Milgram's Reality-Virtuality Continuum," *International Journal of Social Robotics, 1,1. 83-93.*

[65] Kidd, C. and Breazeal, C. (2004). "Effect of a Robot on User Perceptions." *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), Sendai, Japan. Volume 4, 3559-3564.

[66] Bainbridge, W., Hart, J., Kim, E., & Scassellati, B. (2008) The effect of presence on human-robot interaction. *IEEE International Symposium on Robot and Human Interactive Communication,* Munich, Germany.

[67] Powers, A., Kiesler, S., Torrey, C. & Fussell, S. (2007). Comparing a Computer Agent with a Humanoid Robot, 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2007).

[68] Sören Krach, Frank Hegel, Britta Wrede, Gerhard Sagerer, Ferdinand Binkofski, and Tilo Kircher (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLoS ONE,* 3(7). Published online 2008 July 9.

[69] URL for the Personal Robots Group http://robotic.media.mit.edu

[70] URL for Cynthia Breazeal http://www.media.mit.edu/~cynthiab

[71] Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

[72] Breazeal, C. (2003), "Regulation and entrainment for human-robot interaction," *International Journal of Experimental Robotics*, 21(10-11), pp. 883-902.

[73] Breazeal, C., Kidd, C., Thomaz, A.L., Hoffman, G., Berlin, M., "Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork." in Proc. of the International Conference on Intelligent Robotics and Systems, 2005.

[74] Breazeal, C. and Scassellati. B. (2000), "Infant-like social interactions between a robot and a human caregiver," in K. Dautenhahn (ed), *Adaptive Behavior*, 8:1, 47—72.

[75] Breazeal, C. (2003), "Emotion and sociable humanoid robots," E. Hudlika (ed), *International Journal of Human Computer Interaction*, 59, pp. 119-155.

[76] Breazeal, C. (2004), "Function Meets Style: Insights from Emotion Theory Applied to HRI," *IEEE Transactions in Systems, Man, and Cybernetics*, *Part C*, 34(2), 187-194.

[77] C. Breazeal and B. Scassellati (1999). "A context-dependent attention system for a social robot," in

Proceedings of the Sixteenth International Joint Conference on Artifical Intelligence (IJCAI 99). Stockholm, Sweden, 1146--1151.

[78] Breazeal, C., Edsinger, A., Fitzpatrick, P., Scassellati, B., and Varchavskaia, P. (2000), "Social Constraints on Animate Vision," in *IEEE Intelligent Systems, Special Issue on Humanoid Robotics*, 15:4, 32—37.

[79] Thomaz, A. L., Berlin, M. and Breazeal, C. (2005), "An Embodied Computational Model of Social Referencing." *Proceedings of 14th IEEE Workshop on Robot and Human Interactive Communication*, Nashville, TN.

[80] Breazeal, C. and Thomaz, A. (2008) Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers. *Connection Science, 20(2-3), 91-100.*

[81] Thomaz, A. L., & Breazeal, C. (2008). Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence (AIJ), 172(6-7), 716-737.*

[82] Berlin, M., Gray, J., Thomaz, A. L., & Breazeal, C. (2006). Perspective taking: An organizing principle for learning in human-robot interaction. *Proceedings of the 21st National Conference on Artificial Intelligence* (AAAI-06). Boston, MA.

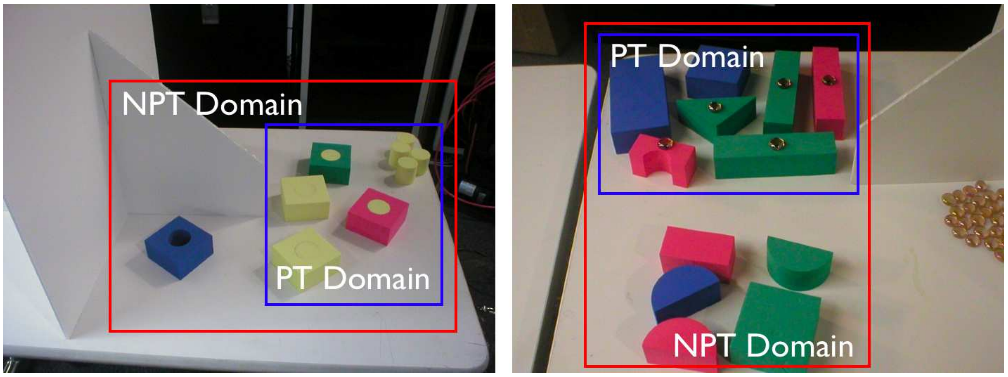
[83] Breazeal, C., & Berlin, M. (2008). Spatial scaffolding for sociable robot learning. *Proceedings of the 23rd Conference on Artificial Intelligence* (AAAI-08). Chicago, IL.

[84] M. Berlin, C. Breazeal, C. Chao (2008). "Spatial Scaffolding Cues for Interactive Robot Learning". Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008). Nice, France.

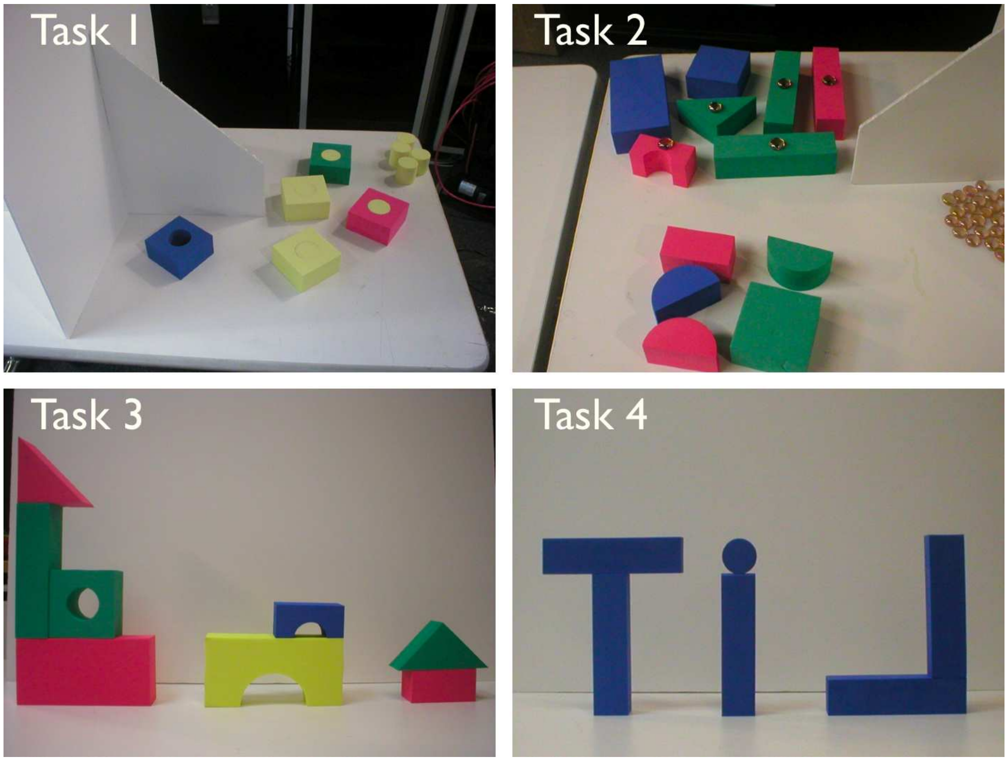
[85] C. Breazeal, G. Hoffman and A. Lockerd (2004). "Teaching and Working with Robots as a Collaboration." In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS). 1030-1037.*

[86] Breazeal, C., & Thomaz, A. L.(2008). Learning from human teachers with socially guided exploration. *Proceedings of the 2008 IEEE International Conference on Robotics and Automation* (ICRA-08). Pasadena, CA.

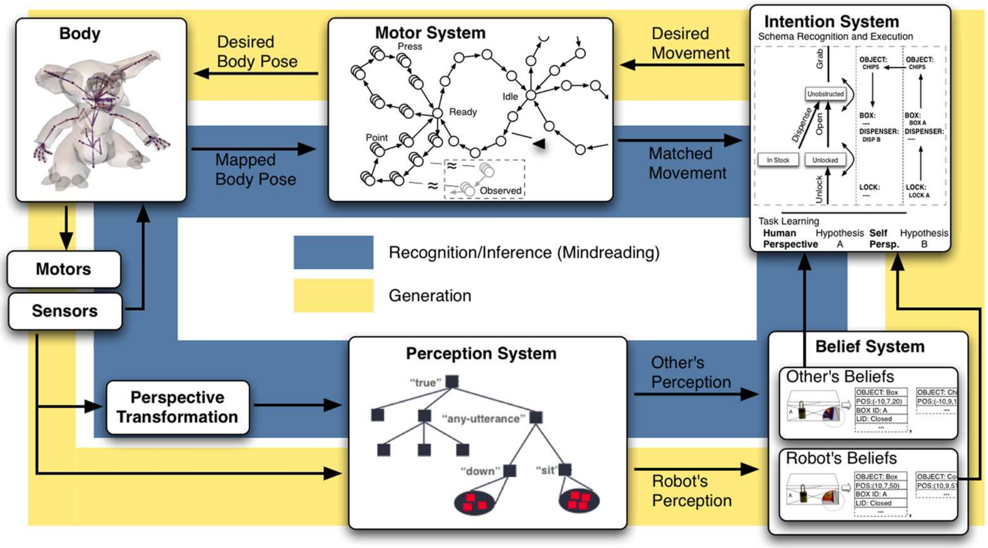
[87] C. Breazeal and L. Aryananda (2002), "Recognizing affective intent in robot directed speech", *Autonomous Robots*, 12(1), pp. 85--104.



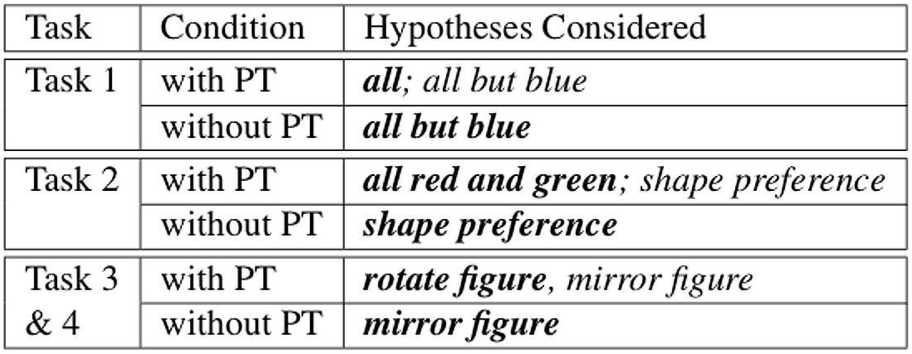
241x88mm (150 x 150 DPI)



241x181mm (150 x 150 DPI)

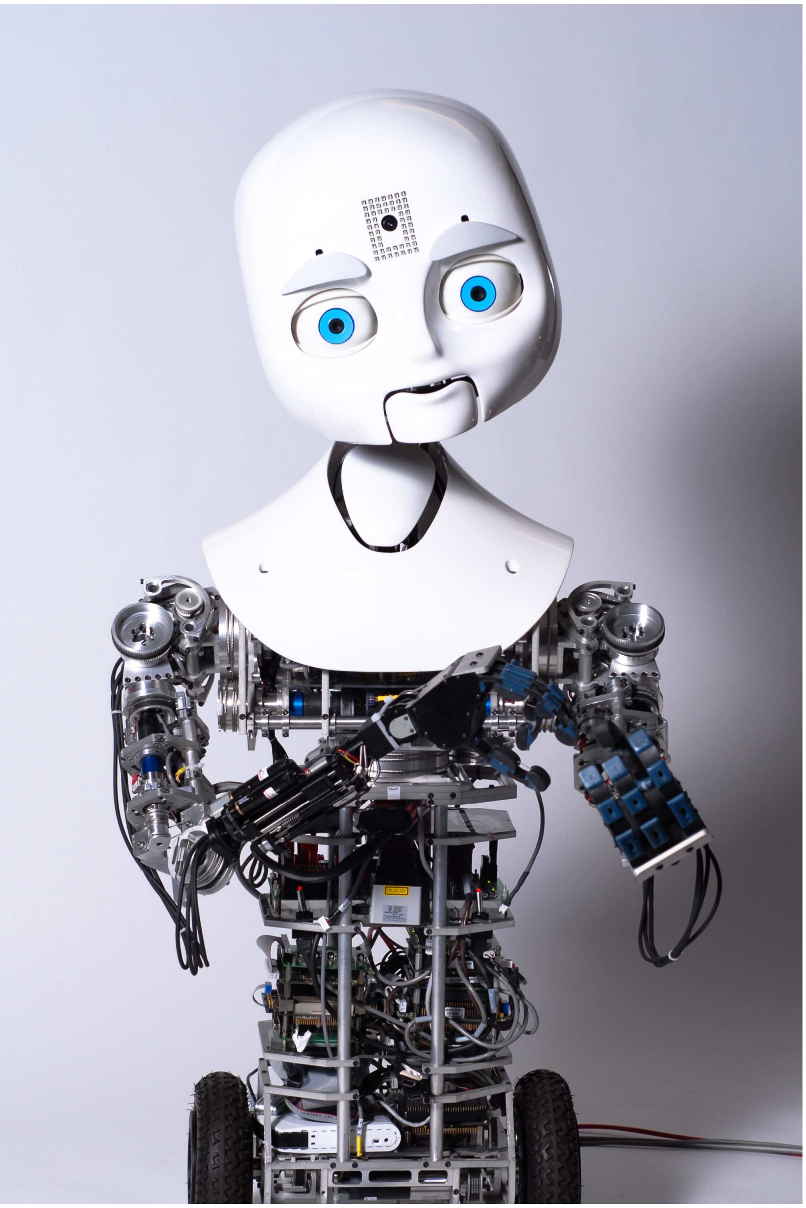


135x76mm (300 x 300 DPI)



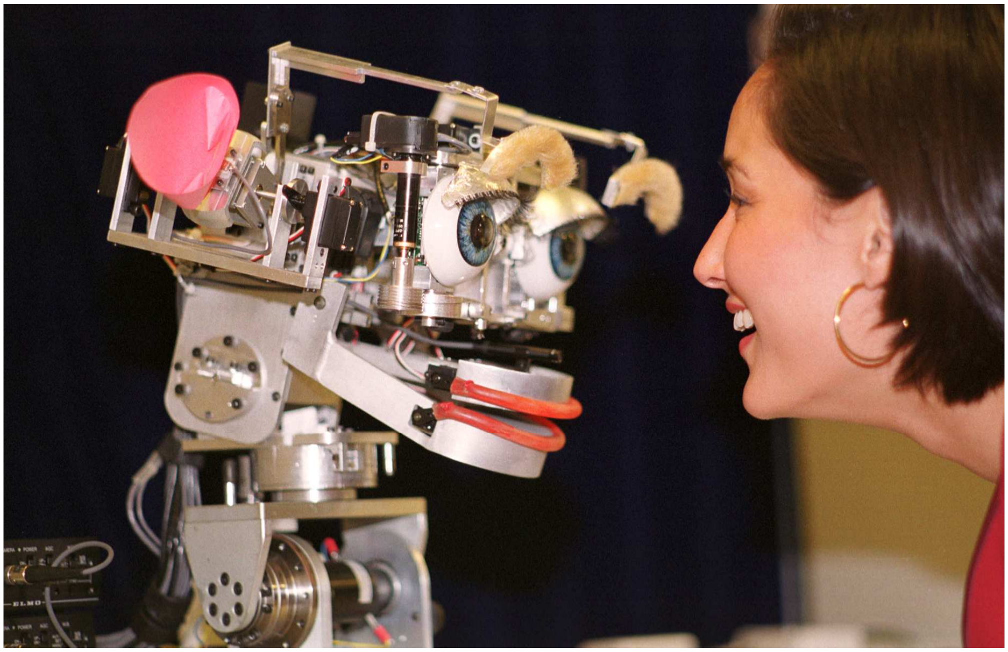
Task	Condition	Hypotheses Considered
Task 1	with PT	***all***; *all but blue*
	without PT	***all but blue***
Task 2	with PT	***all red and green***; *shape preference*
	without PT	***shape preference***
Task 3 & 4	with PT	***rotate figure***, *mirror figure*
	without PT	***mirror figure***

101x44mm (300 x 300 DPI)

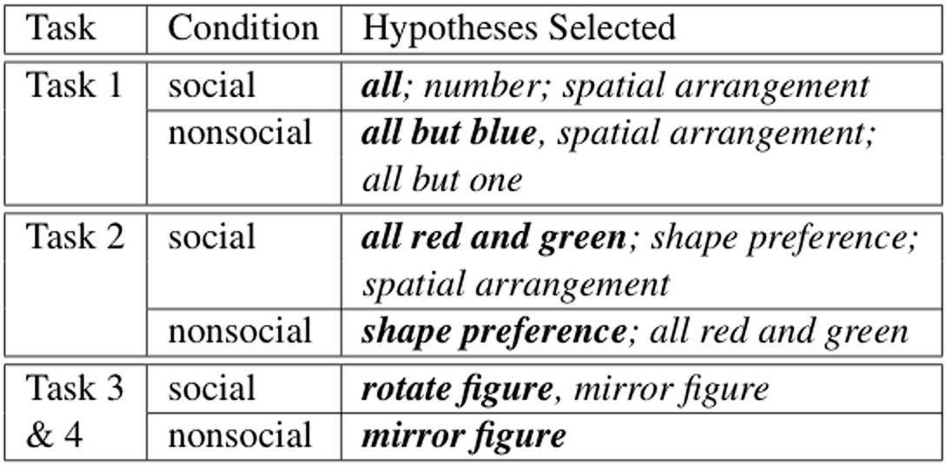


169x253mm (300 x 300 DPI)

254x254mm (100 x 100 DPI)

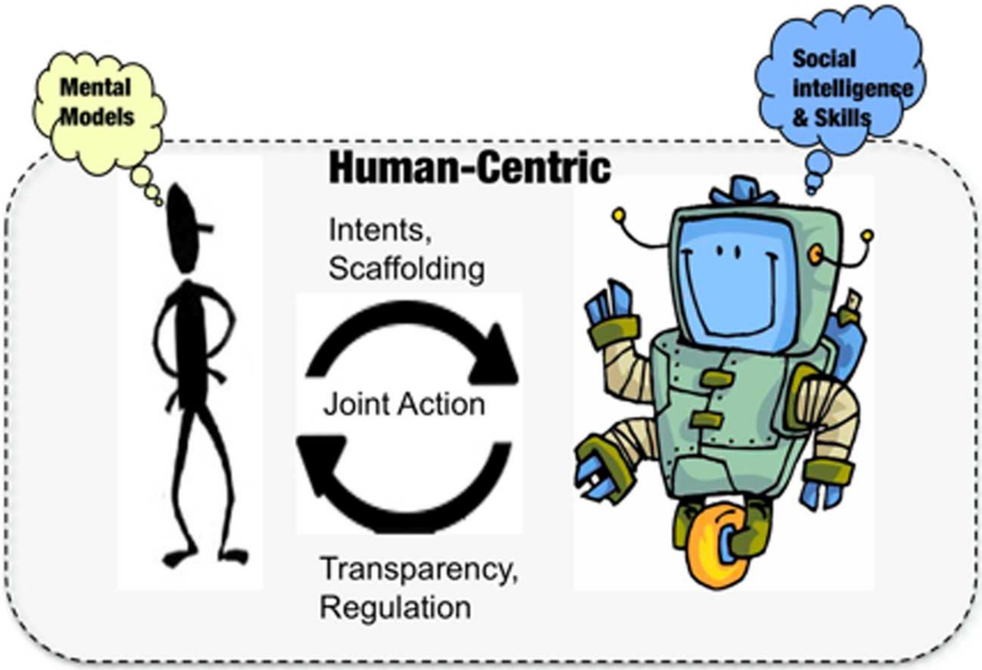


78x50mm (600 x 600 DPI)

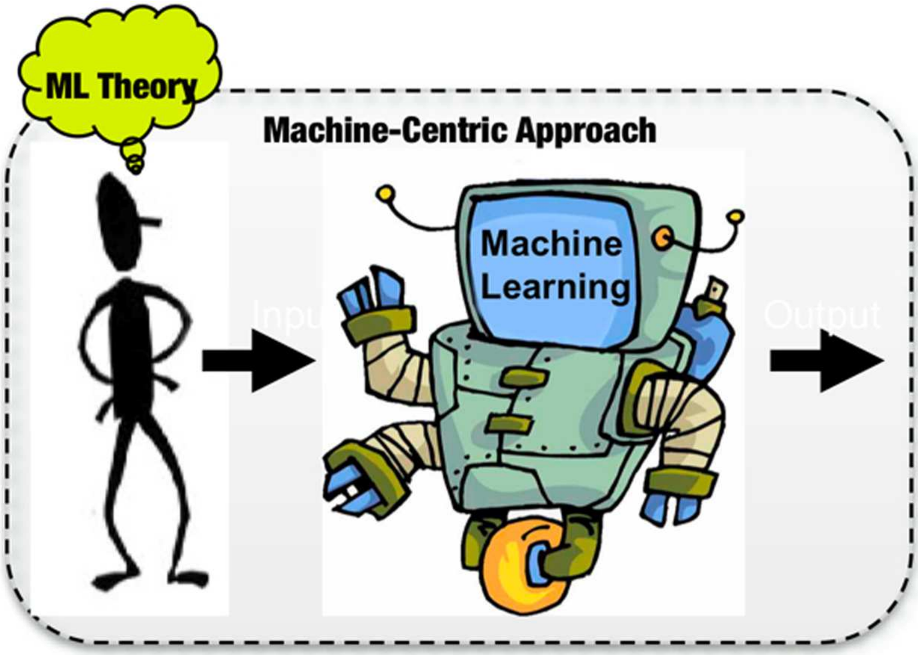


Task	Condition	Hypotheses Selected
Task 1	social	***all***; *number; spatial arrangement*
	nonsocial	***all but blue***, *spatial arrangement; all but one*
Task 2	social	***all red and green***; *shape preference; spatial arrangement*
	nonsocial	***shape preference***; *all red and green*
Task 3 & 4	social	***rotate figure***, *mirror figure*
	nonsocial	***mirror figure***

97x49mm (300 x 300 DPI)



112x76mm (300 x 300 DPI)



113x76mm (300 x 300 DPI)