A Survey of Socially Interactive Robots:
Concepts, Design, and Applications

Terrence Fong [a,b], Illah Nourbakhsh [a], and Kerstin Dautenhahn [c]

[a] The Robotics Institute, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

[b] Institut de production et robotique, Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

[c] Department of Computer Science, The University of Hertfordshire
College Lane, Hatfield, Hertfordshire AL10 9AB, United Kingdom

---

Abstract

This report reviews "socially interactive robots": robots for which social human-robot interaction is important. We begin by discussing the context for socially interactive robots, emphasizing the relationship to other research fields and the different forms of "social robots". We then present a taxonomy of design methods and system components used to build socially interactive robots. Following this taxonomy, we survey the current state of the art, categorized by use and application area. Finally, we describe the impact of these these robots on humans and discuss open issues.

Key words: Human-robot interaction, sociable robot, social robot, socially interactive robot

---

1 Introduction

1.1 The history of social robots

From the beginning of biologically inspired robots, researchers have been fascinated by the possibility of interaction between a robot and its environment, and by the possibility of robots "behaving socially", i.e. interacting with each other. Figure 1 shows the robotic tortoises built by Walter in the late 1940's[104]. By means of headlamps attached to the robots front, and by using positive phototaxis, the two robots interacted in a seemingly "social" manner, but without explicit communication or mutual recognition.

As the field of artificial life emerged, researchers began applying principles such as stigmergy (indirect communication between individuals via modifications made to the shared environment) to achieve "collective" or "swarm" robot behavior. Stigmergy was first described by Grassé as a way to explain how social insect societies can collectively produce complex behavior patterns and physical structures, whilst each individual seemingly works on its own[19].

Deneubourg and his collaborators pioneered the first experiments on stigmergy in simulated and physical "ant-like robots" [75,13] in the early 1990's. Since then, numerous researchers have developed robot collectives[130,150] and have used robots as models for studying social insect behavior[129].

Similar principles can be found in multi-robot or

Email addresses: `terry@ri.cmu.edu` (Terrence Fong), `illah@ri.cmu.edu` (Illah Nourbakhsh), `K.Dautenhahn@herts.ac.uk` (Kerstin Dautenhahn).

Fig. 1. Precursors of social robotics: Walters tortoises, Elmer and Elsie, “dancing” around each other.

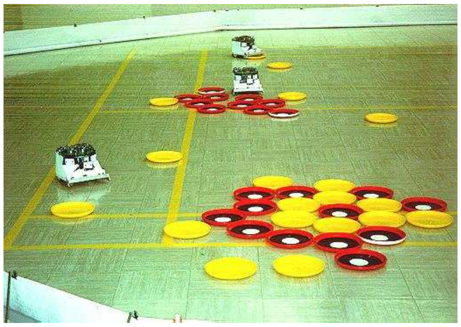

Fig. 2. U-Bots sorting objects[150]

distributed robotic systems research[144]. Some of the interaction mechanisms employed are communication[9], interference among agents[98], and aggressive competition[231]. Common to these “group-oriented”‘social robots approaches is maximizing benefit (e.g., task performance) through collective action (Figures 2, 3, and 4).

The research described thus far uses principles of self-organization and behavior that are inspired by social insect societies. Such societies are anonymous, homogeneous groups in which individuals do not matter. This type of “social behavior” has proven to be an attractive model for robotics, particularly because it enables groups of relatively simple robots perform difficult tasks (e.g., playing soccer).

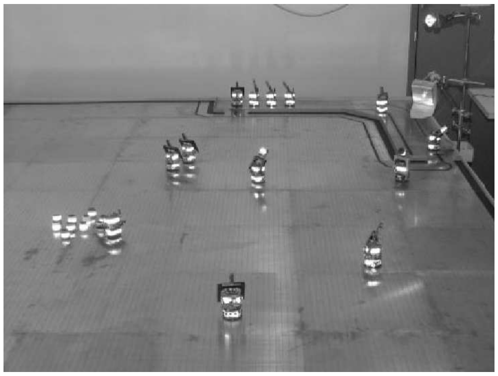

Fig. 3. Khepera robots foraging for “food” [129]

Fig. 4. Collective box-pushing (from[130])

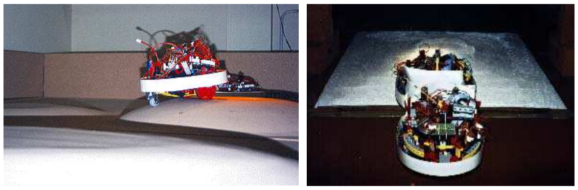

Fig. 5. Early “individual” social robots: “getting to know each other” (left, [56]) and learning by imitation (right, [15,16])

However, many species of mammals (including humans), birds, and other animals often form an individualized society. Individualized societies differ from anonymous societies because the individual matters. Although individuals may live in groups, they form relationships and social networks, they create alliances, and they often adhere to societal norms and conventions[56] (Figure 5).

In [63], Dautenhahn and Billard proposed the following definition:

> Social robots are embodied agents that are part of a heterogeneous group: a society of robots or

> humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other.

Developing such "individual social" robots requires the use of models and techniques different from "group social" collective robots (Figure 6). In particular, social learning and imitation, gesture and natural language communication, emotion, and recognition of interaction partners are all important factors. Moreover, most research in this area has focused on the application of "benign" social behavior. Thus, social robots are usually designed as assistants, companions, or pets, in addition to the more traditional role of servants.

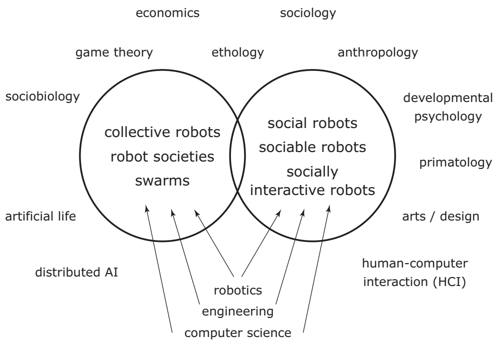

Fig. 6. Fields of major impact. Note that "collective robots" and "social robots" overlap where individuality plays a lesser role.

1.2 Social robots and social embeddedness: concepts and definitions

Robots in individualized societies exhibit a wide range of social behavior, regardless if the society contains other social robots, humans, or both. In [28], Breazeal defines four classes of social robots in terms of (1) how well the robot can support the social model that is abscribed to it and (2) the complexity of the interaction scenario that can be supported:

Socially evocative. Robots that rely on the human tendency to anthropomorphize and capitalize on feelings evoked when humans nurture, care, or involved with their "creation".

Social interface. Robots that provide a "natural" interface by employing human-like social cues and communication modalities. Social behavior is only modeled at the interface, which usually results in shallow models of social cognition.

Socially receptive. Robots that are socially passive but that can benefit from interaction (e.g. learning skills by imitation). Deeper models of human social competencies are required than with social interface robots.

Sociable. Robots that pro-actively engage with humans in order to satisfy internal social aims (drives, emotions, etc). These robots require deep models of social cognition.

Complementary to this list we can add the following three classes:

Socially situated. Robots that are surrounded by a social environment that they need to perceive and react to[70]. Socially situated [1] robots must be able to distinguish between other social agents and various objects in the environment.

Socially embedded. Robots that are: (a) situated in a social environment and interact with other agents and humans; (b) structurally coupled with their social environment; and (c) at least partially aware of human interactional structures (e.g., turn-taking)[70].

Socially intelligent. Robots that show aspects of human style social intelligence, based on deep models of human cognition and social competence[58], i.e. some correspondence exists between social intelligence in humans and artificial social intelligence in robots [56].

---

[1] Other researchers place different emphasis on what *socially situated* implies (e.g., [140]).

1.3 Socially interactive robots

For the purposes of this paper, we use the term “socially interactive robots” to describe robots for which social interaction plays a key role. We do this, not to introduce another class of social robot, but rather to distinguish these robots from other robots that involve “conventional” human-robot interaction, such as those used in teleoperation scenarios.

In this paper, we focus specifically on peer-to-peer, human-robot interaction. However, generally speaking, the term *social robots* also applies to groups of robots that behave socially without human contact. In addition, socially interactive robots show some of the following characteristics of human-human interaction competencies:

- express and/or perceive emotions
- communicate with high-level dialogue
- learn/recognize models of other agents
- establish/maintain social relationships
- use natural cues (gaze, gestures, etc.)
- exhibit distinctive personality and character
- may learn/develop social competencies

Socially interactive robots can be used for a variety of purposes: as research platforms, as toys, as educational tools, or as therapeutic aids. The common, underlying assumption is that humans prefer to interact with machines in the same way that they interact with other people. A survey and taxonomy of current applications is given in Section 3.

Socially interactive robots operate as partners, peers or assistants, which means that they need to exhibit a certain degree of adaptability and flexibility to drive the interaction with a wide range of humans. Socially interactive robots can have different shapes and functions, ranging from robots whose sole purpose and only task is to engage people in social interactions (Kismet, Cog, etc.) to robots that are engineered to adhere to social norms in order to fulfill a range of tasks in human-inhabited environments (Pearl, Sage, etc.)[27,170,186,203].

Thus, some robots pro-actively encourage social interaction while others show their social competence only in reaction to human behavior. Some of these robots might use deep models of human interaction while others might rely on humans attributing mental states and emotions to the robot role[57,66,78,182].

Designing socially interactive robots means considering the human in the loop: as interaction partner, as designer, and as observer. Thus, robots can be social if they affect the human perception of the robot as a social entity. Nevertheless, the focus of socially interactive robots is on the interaction between human and robot, rather than on any internal characteristics of the robot itself.

1.4 Why socially interactive robots?

The domains that we are interested in are those in which robots must act as social interaction partners. In such applications, robots must exhibit social interaction skills, either because such skills are required for solving specific tasks, or because the primary function of the robot is to interact socially with people. A discussion of application domains, design spaces, and desirable social skills for robots is given in [61] and [62].

One area where social interaction is desirable is that of “robot as persuasive machine”[83], i.e., the robot is used to change the behavior, feelings or attitudes of humans. This is the case when robots mediate human-human interaction, as in autism therapy[238]. Another area is “robot as avatar”[180], in which the robot functions as a representation of, or representative for, the human. For example, if a robot is used for remote communication, it may need to act socially in order to effectively convey information.

In certain scenarios, it may be desirable for a robot to develop its interaction skills over time. For example, a pet robot that accompanies a child through his childhood may need to improve its skills in order to maintain interest. Learned development of social (and other) skills is a primary concern of epigenetic robotics[246,63].

Some researchers design socially interactive robots simply to study embodied models of social behavior. For this use, the challenge is to build robots that have an intrinsic notion of sociality, that develop social skills and bond with people, and that can show empathy and true understanding. At present, such robots remain a distant goal[57,63], the achievement of which will require contributions from other research areas such as artificial life, developmental psychology and sociology[194].

Although socially interactive robots have already been used with success, much work remains to increase their effectiveness. For example, in order for socially interactive robots to be accepted as "natural" interaction partners, they need more sophisticated social skills, such as the ability to recognize social context and convention.

Additionally, socially interactive robots will eventually need to support a wide range of users: different genders, different cultural and social background, different ages, etc. In many current applications, social robots only engage in short-term interaction (e.g., a museum tour) and can afford to treat all humans in the same manner. But, as soon as a robot becomes part of a person's life, that robot will need to be able to treat him as a distinct individuals [58].

In the following, we closely examine the concepts raised in this introductory section. We begin by describing different design methods. Then, we present a taxonomy of system components, focusing on the design issues unique to socially interactive robots. We conclude by discussing open issues and core challenges.

2 Methodology

2.1 Design approaches

Humans are experts in social interaction. Thus, if technology adheres to human social expectations, people will find the interaction enjoyable, feeling empowered and competent[191]. Many researchers, therefore, explore the design space of anthropomorphic (or zoomorphic) robots, trying to endow their creations with characteristics of intentional agents. For this reason, more and more robots are being equipped with faces, speech recognition, lip-reading skills, and other features and capacities that make robot-human interaction more "human-like" or at least "creature-like"[60,70].

From a design perspective, we can classify how socially interactive robots are built in two primary ways. With the first approach, "biologically inspired", designers try to create robots that internally simulate, or mimic, the social intelligence found in living creatures. With the second approach, "functionally designed", the goal is to construct a robot that appears outwardly to be socially intelligent, even if the internal design does not have a basis in science.

Robots have limited perceptual, cognitive, and behavioral abilities compared to humans. Thus, for the foreseeable future, there will continue to be significant imbalance in social sophistication between human and robot[29]. As with expert systems, however, it is possible that robots may become highly sophisticated in restricted areas of socialization, e.g., infant-caretaker relations.

Finally, differences in design methodology means that the evaluation and success criteria are almost always different for different robots. Thus, it is hard to compare socially interactive robots outside of their target environment and use.

2.1.1 Biologically inspired

With the "biologically inspired" approach, designers try to create robots that internally simulate, or mimic, the social behavior or intelligence found in living creatures. Biologically inspired designs are based on theories drawn from natural and social sciences, including anthropology, cognitive science, developmental psychology, ethology, sociology, structure of interaction, and theory of mind. Generally speaking, these theories are used to guide the design of robot cognitive, behavioral, motivational (drives and emotions), motor and perceptual systems.

There are two arguments made for drawing inspiration from biological systems. First, some researchers contend that nature is the best model for "life-like" activity. Specifically, they hypothesize that for a robot to be understandable by humans, it must have a naturalistic embodiment, it must interact with its environment in the same way living creatures do, and it must perceive the same things that humans find to be salient and relevant[246].

The second rationale for biologically inspiration is that it allows us to directly examine, test and refine those scientific theories upon which the design is based[1]. This is particularly true with humanoid robots. Cog, for example, is a general-purpose humanoid platform intended for exploring theories and models of intelligent behavior and learning [203].

Some of the theories commonly used in biologically inspired design are:

Ethology. Ethology refers to the observational study of animals in their natural setting[137]. Ethology can serve as a basis for design because it describes the types of activity (comfort-seeking, play, etc.) a robot needs to exhibit in order to appear life-like [6]. Ethology is also useful for addressing a range of behavioral issues such as concurrency, motivation, and instinct.

Structure of interaction. Analysis of interactional structure (such as instruction, cooperation, etc.) can help focus design of perception and cognition systems by identifying key interaction patterns [238]. Dautenhahn, Ogden and Quick use explicit representations of interactional structure to design "interaction aware" robots[70]. Dialogue models, such as turn-taking in conversation, can also be used in design as in [147].

Theory of mind. Theory of mind refers to those social skills that allow humans to correctly attribute beliefs, goals, perceptions, feelings, and desires to the self and others[239]. One of the critical precursors to these skills is joint (or shared) attention: the ability to selectively attend to an object of mutual interest [10]. Joint attention can aid design, by providing guidelines for recognizing and producing social behaviors such as gaze direction, pointing gestures, etc.[33,203].

Developmental psychology. Developmental psychology has been cited as effective mechanism for creating robots that can engage in natural social exchanges. As an example, the design of Kismet's "synthetic nervous system", particularly the perceptual and behavioral aspects, is heavily inspired by the social development of human infants[27]. Additionally, theories of child cognitive development, such as Vygotsky's "child in society" [134], can offer a framework for constructing robot architecture and social interaction design [63,64].

2.1.2 Functionally designed

With the "functionally designed" approach, the objective is to design a robot that appears outwardly to be socially intelligent, even if the internal design does not have a basis in science or nature. This approach assumes a deep understanding of how the mind really works is not necessary for creating the impression of an agent driven by beliefs and desires. It is sufficient to describe the mechanisms (sensations, traits, folk-psychology, etc.) by which people in everyday life understand socially intelligent creatures[182].

In contrast to their "biologically inspired" counterparts, functionally designed robots generally have constrained operational and performance objectives. Consequently, these "engineered" robots may need only to generate certain effects and experiences with the user, rather than having to withstand extensive scrutiny for "life-like" capabilities.

Some motivations for "functional design" are:

- The robot may only need to be superficially socially competent. This is particularly true when only short-term interaction or limited quality of interaction is required.
- The robot may have limited embodiment, capability for interaction, or may be constrained by the environment.
- Even limited social expression can help improve the affordances and usability of a robot. In some applications, recorded or scripted speech may be sufficient for human-robot dialogue.

- Artificial designs can provide compelling interaction. Many video games and electronic toys fully engage and occupy attention, even if the artifacts do not have real-world counterparts.

The three techniques most often used in functional design are:

Human-Computer Interaction (HCI) design. Robots are increasingly being developed using HCI techniques, including cognitive modeling, contextual inquiry, heuristic evaluation, and empirical user testing[167]. User studies are conducted, often throughout development, in order to understand the user's activities and to assess the interface (or system) usability.

Nourbakhsh describes the design of a personal rover, which is intended to be a creative and expressive tool, for non-specialists such as children [82]. The design of this rover is guided by an experience design document developed through a user study. Scheeff et al. discusses the development of Sparky, a telerobotic "creature" built to explore non-conventional HCI[205]. Sparky's design was inspired both by principles of traditional animation and cartooning (e.g., how to evoke emotional state through motion) as well as heuristic design goals (exhibit smooth motion, make the body active, etc.).

Systems Engineering. Systems engineering involves the top-down development of a systems' functional and physical requirements from a basic set of objectives. The purpose is to organize information and knowledge to facilitate and control the planning, development, and operation of the system[197]. Systems engineering is frequently used in robot development, although structured design techniques (configuration control, work breakdown structure, etc.) tend to be applied informally.

A basic characteristic of system engineering is that it only places emphasis on the design of critical-path system elements. Pineau et al., for example, describe mobile robots designed to assist the elderly in daily living[186]. Because these robots operate in a highly structured domain their design centers on a collection of task-based behaviors: autonomous navigation, speech recognition, face tracking, etc. rather than broad social interaction.

Iterative design. Iterative (or sequential) design, is the process of revising a design through a series of test and redesign cycles[215]. It is typically used to address design failures or to make improvements based on information from evaluation or use. Iterative design can be an effective design technique, particularly when a system or its target environment are difficult to model analytically. However, because suggestions for improvement are often based on anecdotal data, design changes may result in little or no improvement.

Willeke et al. describe a series of museum robots, each of which was designed based on lessons learned from preceding generations[241]. The design objective was to attract people to interact, based on environmental constraints and using simple interaction models. Schulte et al. discuss design for short-term and spontaneous interaction between Minerva[228], another tour guide robot, and crowds of people[206].

2.2 Design issues

All robot systems, whether socially interactive or not, must address a number of common design problems. These include cognition (planning, decision making), perception (navigation, environment sensing), action (mobility, manipulation), human-robot interaction (user interface, input devices, feedback display) and architecture (control, electromechanical, system). Socially interactive robots, however, must also address those issues imposed by social interaction[27,58]. In particular:

Human-oriented perception. A socially interactive robot must proficiently perceive and interpret human activity and behavior. This includes detecting and recognizing gestures, monitoring and classifying activity, discerning intent and social cues, and measuring the human's feedback.

Natural human-robot interaction. Humans and robots should communicate as peers who know each other well, such as musicians playing a duet [208]. To achieve this, the robot must manifest believable behavior: it must establish appropriate social expectations, it must regulate social interaction (using dialogue and action), and it must follow social convention and norms.

Readable social cues. A socially interactive robot must send signals to the human in order to: (1) provide feedback of its internal state; and (2) allow human to interact in a facile, transparent manner. Because robots are constructed, they have limited channels and capacity for emotional expression. These include facial expression, body and pointer gesturing, and vocalization (both speech and sound).

Real-time performance. Socially interactive robots must operate at human interaction rates. Thus, a robot needs to simultaenously exhibit competent behavior, convey attention and intentionality, and handle social interaction.

In the following sections, we review design issues that are unique to socially interactive robots. Although we do not discuss every aspect of design (e.g., architecture), we feel that addressing each of the following is critical to building an effective social robot.

2.2.1 Embodiment

Biological bodies have evolved in order to adapt to specific internal and environmental constraints. This "evolved embodiment" plays a significant role in the emergence of cognition and emotion[183]. In particular, the design of effectors and sensors is often tightly coupled to how neural processing is performed. For example, the visual axes in a fly's eye have non-uniform sampling, which enables simple processing for motion parallax.

We define embodiment as "that which establishes a basis for structural coupling by creating the potential for mutual perturbation between system and environment"[70]. Thus, embodiment is grounded in the relationship between a system and its environment. The more a robot can perturb an environment, and be perturbed by it, the more it is embodied[59]. This also means that social robots do not necessarily need a *physical* body. For example, conversational agents[47] might be embodied to the same extent as robots with limited acuation.

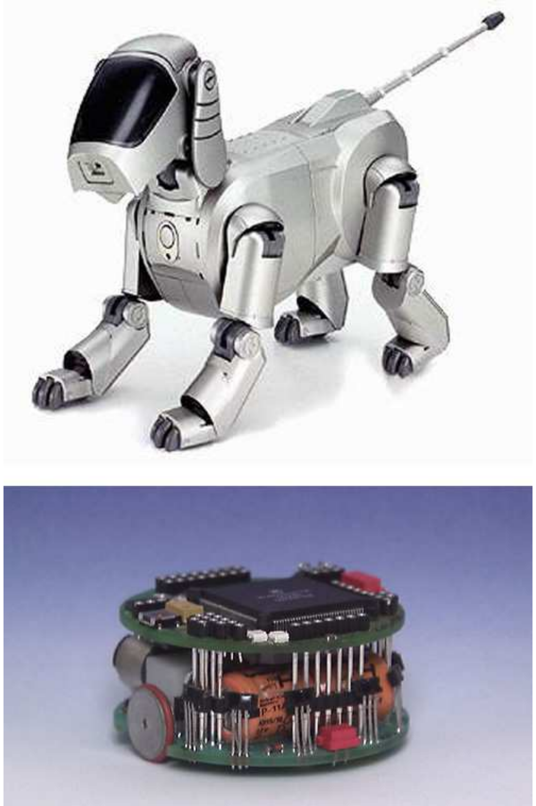

Fig. 7. Sony Aibo ERS-110 (top) and K-Team Khepera (bottom)

An important benefit of this "relational definition" is that it provides an opportunity to quantify embodiment. For example, one might measure embodiment in terms of the complexity of the relationship between robot and environment over all possible interactions (i.e., all perturbatory channels).

All robots are embodied, but some are more embodied than others[70]. Consider the difference between Aibo (Sony) and Khepera (K-Team), as shown in Figure 7. Aibo has approximately 20 actuators (joints across mouth, heads, ears, tails, and legs) and a variety of sensors (touch, sound, vision and proprioceptive). In contrast, Khepera has 2 actuators (independent wheel control) and an array of infrared proximity sensors. Because Aibo has more perturbatory channels and bandwidth at its disposal than does Khepera, it can be considered to be more strongly embodied than Khepera.

Morphology

The form and structure of a robot is important because it helps establish social expectations. Physical appearance biases interaction. A robot that resembles a dog will be treated differently (at least initially) than one which is anthropomorphic. Moreover, the relative familiarity (or strangeness) of a robot's morphology can have profound effects on its accessibility, desirability, and expressiveness.

The choice of a given form may also constrain the human's ability to interact with the robot. For example, Kismet has a highly expressive face. But because it is designed as a head, Kismet is unable to interact when touch (e.g., manipulation) or displacement (self movement) is required.

To date, most research in human-robot interaction has not explicitly focused on design, at least not in the traditional sense of industrial design. Although knowledge from other areas of design (including product, interaction and stylized design) can inform robot construction, much research remains to be performed.

Design considerations

When designing a robot's form and structure, there are a number of considerations that should be taken in to account [76]. First, a robot's morphology must match its intended function. If a robot is designed to perform tasks for the human, then its form must convey an amount of "product-ness" so that the user will feel comfortable using the robot. Similarly, if peer interaction is important, the robot must project an amount of "human-ness" so that the user will feel comfortable in socially engaging the robot.

At the same time, however, a robot's design needs to reflect an amount of "robot-ness". This is needed so that the user does not develop detrimentally false expectations of the robot's capabilities [78].

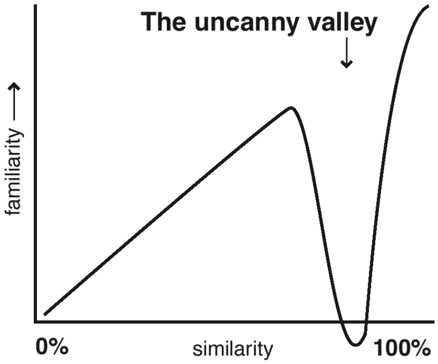

Fig. 8. Mori's "Uncanny Valley" (from [76]).

Finally, if a robot needs to portray a living creature, it is critical that an appropriate degree of familiarity be maintained. Mashiro Mori contends that the progression from a non-realistic to realistic portrayal of a living thing is nonlinear. In particular, there is an "uncanny valley" (see Figure 8) as similarity becomes almost, but not quite perfect. At this point, the subtle imperfections of the recreation become highly disturbing, or even repulsive[192]. Consequently, caricatured representations may be more useful, or effective, than more complex, "realistic" representations.

We classify social robots as being embodied in four broad categories: anthropomorphic, zoomorphic, caricatured, and functional.

Anthropomorphic

Anthropomorphism, from the Greek "anthropos" for man and "morphe" for form/structure, is the tendency to attribute human characteristics to objects with a view to helping rationalize their actions[78]. Anthropomorphic paradigms have widely been used to augment the functional and behavioral characteristics of social robots (see Section 3.5).

Having a naturalistic embodiment is often cited as necessary for meaningful social interaction[122,27,203]. In part, the argument is that in order for a robot to interact with humans as hu-

Fig. 9. Leonardo has a “creature-like” appearance (courtesy Cynthia Breazeal, MIT Media Lab)

mans do (through gaze, gesture, vocalization, etc.), it must be structurally and functionally similar to a human. Moreover, if a robot is to learn from humans (e.g., through imitation), then it should be capable of behaving similarly to humans [14].

The role of anthropomorphism is to function as a mechanism (for design, for interpreting behavior, etc.) through which social interaction can be facilitated. Thus, the ideal use of anthropomorphism is to present an appropriate balance of illusion (to lead the user to believe that the robot is sophisticated in areas where the user will not encounter its failings) and functionality (to provide capabilities necessary for supporting human-like interaction) [76,116].

Zoomorphic

An increasing number of entertainment, personal, and toy robots have been designed to imitate living creatures (see Section 3.3). For these robots, a zoomorphic embodiment is important for establishing human-creature relationships.

The most common designs are inspired by household animals, such as the Sony Aibo (Figures 7 and 22) dog and the Omron NeCoRo (Figure 25), with the objective of creating robotic “companions”. Other designs, such as Leonardo (Figure 9), have “‘creature-like” appearance but do not have real-world counterparts. Avoiding the “uncanny valley” may be easier with zoomorphic design because human-creature relationships (e.g., owner-pet) are often simpler than human-human relationships. Thus, our expectations of what constitutes “realistic” and “unrealistic” animal morphology tends to be lower.

Caricatured

Animators have long demonstrated that a character does not have to appear realistic in order to be believable[226]. Caricature, for example, exaggerates distinctive or striking features in order to produce comic effect. Moreover, simplified or stereotypical representations, such as cartooning, can be used to create desired interaction biases (e.g., implied abilities) and to focus attention on, or distract attention from, specific robot features.

Scheeff et al. discusses how techniques from traditional animation and cartooning can be used in social robot design[205]. Schulte et al. describe how a caricatured human face (two eyes with eyebrows and a mouth) can provide an explicit “focal point” on which people can focus their attention[206]. Similarly, Severinson-Eklund et al. describe the use of a small mechanical character, CERO, as a robot “representative” (see Figure 10)[207].

Functional

Some researchers argue that a robot’s embodiment should first, and foremost, reflect the tasks it must perform. The choice and design of physical features is thus guided purely by operational objectives. This type of embodiment appears most often with functionally designed robots, especially service robots.

Health care robots, for example, may be required to assist elderly, or disabled, patients in moving about. Thus, features such as handle bars and cargo space, may need to be part of the design [186].

The design of toy robots also tends to reflect functional requirements. Toys must minimize production cost, be appealing to children, and be capable of facing the wide variety of situations that can experienced during play [152].

Fig. 10. CERO (from Severinson-Eklund 2002)

2.2.2 Emotion

Emotions play a significant role in human behavior, communication and social interaction[7,114]. Emotions influence cognitive processes, particularly problem solving and decision making. Emotions also guide action, control resource usage, and shape dialogue. Emotions are complex phenomena and often tightly coupled to social context. Moreover, much of emotion is physiological and depends on embodiment[86,179,183].

Three primary theories are used to describe emotions. The first approach describes emotions in terms of discrete categories (e.g., happiness). A good review of "basic emotions" is [80]. The second approach characterizes emotions using continuous scales or basis dimensions, such as *arousal* and *valence*[200]. The third approach, componential theory, acknowledges the importance of both categories and dimensions[188,215].

In recent years, emotion has increasingly been used in interface and robot design, primarily because of the recognition that people tend to treat computers as they treat other people[8,41,43,47,178,191]. Moreover, many studies have been performed to integrate emotions into products including electronic games, toys, and software agents[11].

Artificial Emotions

Artificial emotions are used in social robots for several reasons. The primary purpose, of course, is that emotion helps facilitate believable human-robot interaction[42,172]. Artificial emotion can also provide feedback to the user, such as indicating the robot's internal state, goals and (to an extent) intentions[11,21,123]. Lastly, artificial emotions can act as a control mechanism, driving behavior and reflecting how the robot is affected by, and adapts to, different factors over time [25,40,153,232,233]. Two overviews of emotional control are [44] and [45].

Numerous architectures have been proposed for artificial emotions[27,40,81,106,193,232,233]. Some closely follow emotional theory, particularly in terms of how emotions are defined and generated. Arkin et al. discuss how ethological and componential emotion models are incorporated into Sony's entertainment robots[6]. Cañamero and Fredslund describe an affective activation model that regulates emotions through stimulation levels[42].

Other architectures are only loosely inspired by emotional theory and tend to be designed in an ad-hoc manner. Nourbakhsh et al. detail a fuzzy state machine based system, which was developed through a series of formative evaluation and design cycles[170]. In this system, the state machine controls both emotion expression and robot action selection. Schulte et al. summarize the design of a simple state machine that produces four basic "moods"[206]. In their approach, mood is viewed purely as a mechanism for facilitating the robot's achievement of navigation goals.

In terms of expression, some robots are only capable of displaying emotion in a limited way, such as individually actuated lips or flashing lights (usually LEDs). Other robots have many active degrees of freedom and can thus provide richer movement and gestures. Kismet, for example, has controllable eyebrows, ears, eyeballs, eyelids, a mouth with two lips and a pan/tilt neck[27].

Emotions as control mechanism

Emotion can be used to determine control precedence between different behavior modes, to coordinating planning, and to trigger learning and adaptation, particularly when the environment is unknown or difficult to predict. One approach is to use computational models of emotions that mimic animal survival instincts, such as escape from danger, look for food, etc.[25,40,92,153,233].

Several researchers have investigated the use of emotion in human-robot interaction. Suzuki et al. describe an architecture in which interaction leads to changes in the robot's emotional state and modifications in its actions[222]. Breazeal discusses how emotions influence the operation of Kismet's motivational system and how this affects its interaction with humans[30]. Nourbakhsh et al. discusses how mood changes can trigger different behavior in Sage, a museum tour robot[170].

Speech

Speech is a highly effective method for communicating emotion. The primary parameters that govern the emotional content of speech are loudness, pitch (level, variation, range), and prosody. Murray and Arnott contend that the vocal effects caused by particular emotions are consistent between speakers, with only minor differences[162].

The quality of synthesized speech is significantly poorer than synthesized facial expression and body language [12]. In spite of this shortcoming, it has proved possible to generate emotional speech. Cahn describes a system for mapping emotional quality (e.g., sorrow) onto speech synthesizer settings, including articulation, pitch, and voice quality[39].

To date, emotional speech has been used in few robot systems. In [26], Breazeal describes the design of Kismet's vocalization system, in which expressive utterances are generated by assembling strings of phonemes with pitch accents. In [170], Nourbakhsh et al. describe how emotions influence synthesized speech production in a tour guide robot, e.g., when the robot is frustrated, voice level and pitch are increased.

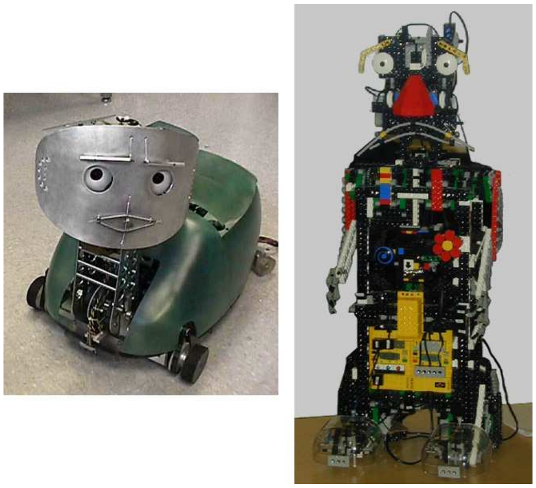

Fig. 11. Actuated faces: Sparky (left) and Feelix (right)

Facial expression

The human face serves many purposes. It displays an individual's motivation, which helps make behavior more predictable and understandable to others. It supplements verbal communication by signaling the speaker's attitude towards the information being spoken. Facial gestures and expressions also communicate information, such as a shrug to express "I don't know", and help regulate dialogue by providing turn-taking cues [27].

The expressive behavior of robotic faces is generally not life-like. This reflects limitations of mechatronic design and control. For example, transitions between expressions tend to be abrupt, occurring suddenly and rapidly, which rarely occurs in nature. The primary facial components used are mouth (lips), cheeks, eyes, eyebrows and forehead. Most robot faces express emotion in accordance with Ekman and Frieser's FACS system[79,214].

Two of the simplest faces (Figure 11) appear on Sparky[205] and Feelix[42]. Sparky's face has 4-DOF (eyebrows, eyelids, and lips) which portray a set of discrete, basic emotions. Feelix is a robot built using the LEGO Mindstorms$^{\text{TM}}$robotic construction kit. Feelix's face also has 4-DOF (two eyebrows, two lips), designed to display six facial expressions (anger, sadness, fear, happiness, surprise, neutral) plus a number of blends.

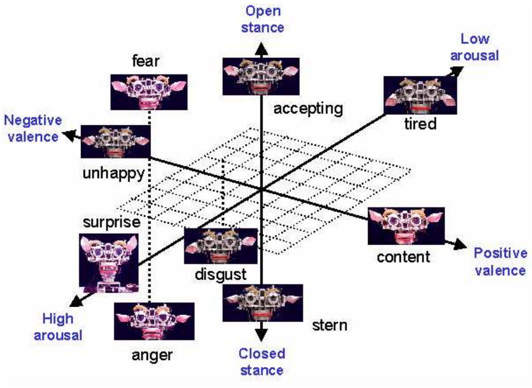

Fig. 12. Various emotions displayed by Kismet

Fig. 13. Saya face robots (Science University of Tokyo)

In contrast to Sparky and Feelix, Kismet's face has fifteen actuators, many of which often work together to display specific emotions (see Figure 12). Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional componential "affect space" (arousal, valence, and stance). Kismet is able to display expressions that map to anger, distrust, fear, happiness, sorrow and surprise[23].

Perhaps the most realistic robot faces are those designed at the Science University of Tokyo[120]. These faces (Figure 13) are explicitly designed to be human-like and incorporate hair, teeth, and a covering silicone skin layer. Numerous control points actuated beneath the "skin" produce a wide range of facial movements and human expression.

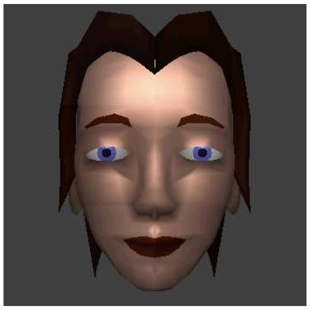

Fig. 14. Vikia has a computer generated face

Instead of using mechanical actuation, another approach to facial expression is to rely on computer graphics and animation techniques[142]. Vikia, for example, has a 3D rendered face of a woman based on Delsarte's code of facial expressions[37]. Because Vikia's face (see Figure 14) is graphically rendered, many degrees of freedom are available for generating expressions.

Body Language

In addition to facial expressions, non-verbal communication is often conveyed through gestures and body movement[12]. Over 90% of gestures occur during speech and provide redundant information[127,149]. To date, most studies on emotional body movement have been qualitative in nature. Frijda, for example, described body movements for a number of basic emotions (Table 1)[87]. Recently, however, some work has begun to focus on implementation issues, such as in [53].

Nakata et al. state that humans have a strong tendency to be cued by motion[163]. In particular, they refer to analysis of dance that shows humans are emotionally affected by body movement. Breazeal and Fitzpatrick contend that humans perceive all motor actions to be semantically rich, whether or not they were actually intended to be[31]. For example, gaze and body direction is generally interpreted as indicating locus of attention.

Table 1
Emotional body movements (adapted from [87]).

Emotion	Body movement
anger	fierce glance; clenched fists; brisk, short motions
fear	bent head, truck and knees; hunched shoulders; forced eye closure or staring
happiness	quick, random movements; smiling;
sadness	depressed mouth corners; weeping
surprise	wide eyes; held breath; open mouth

Mizoguchi et al. discuss the use of gestures and movements to create a sense of familiarity[160]. In their system, they exploit a set of designed expressions, similar to ballet poses, to show emotion through movement. Scheeff et al. describe the design of smooth, natural motions for Sparky[205]. Lim, Ishii and Takanishi describe how walking motions (foot dragging, body bending, etc.) can be used to convey emotions [135].

2.2.3 Dialogue

What is dialogue?

Dialogue is a joint process of communication. It involves sharing of information (data, symbols, context) and control between two (or more) parties [132]. Humans employ a variety of para-linguistic social cues (facial displays, gestures, etc.) to regulate the flow of dialogue [46]. These social cues have also proven to be effective for controlling human-robot dialogue, which is important because the current performance of robot perception (e.g., speech processing) limits the rate at which human-robot dialogue can proceed[24,28].

Dialogue, regardless of form, is meaningful only if it is grounded, i.e., that the symbols used by each party describe common concepts. When symbols differ, information exchange or learning must take place before communication can proceed. Although human-robot communication can occur in many different ways, we consider there to be three significant types of dialogue: low-level (pre-linguistic), non-verbal, and natural language.

Low-level. Billard and Dautenhahn describe a number of experiments in which an autonomous mobile robot was taught a synthetic proto-language[15–17]. Language learning results from multiple spatio-temporal associations across the robot's sensor-actuator state space.

Steels has examined the hypothesis that communication is bootstrapped in a social learning process and that meaning is initially context-dependent[218,219]. In his experiments, a robot dog learns simple words describing the presence of objects (ball, red, etc), its behavior (walk, sit) and its body parts (leg, head).

Non-verbal. There are many non-verbal forms of language, including body positioning, gesturing, and physical action. Since most robots have fairly rudimentary capability to recognize and produce speech, non-verbal dialogue is a useful alternative. Nicolescu and Mataric, for example, describe a robot that asks humans for help, communicating its needs and intentions through its actions[168].

Social conventions, or norms, can also be expressed through non-verbal dialogue. Proxemics, the social use of space, is one such convention [100]. Proxemic norms include knowing how to stand in line, how to enter elevators, how to pass in hallways, etc. Respecting these spatial conventions may involve consideration of numerous factors (administrative, cultural, etc.)[164].

Natural language. Natural language dialogue is determined by a set of factors ranging from the physical and perceptual capabilities of the participants, to the social and cultural features of the situation in which the dialogue is carried out To what extent natural language dialogue interfaces for robots should be based on human-human dialogue is clearly an open issue[207]. Moreover, creating a robot that communicates at a human peer level using natural language remains a grand challenge.

Severinson-Eklundh et al. discuss how users need explicit feedback when interacting with service robots[207]. In particular, they contend that the user always needs to understand what instructions the robot has received and is about to carry out. One way of achieving this is to emphasize providing conversational feedback (e.g., through natural language) through dialogue design.

Fong, Thorpe, and Baur describe how high-level dialogue can enable a human to provide assistance to a robot[84,85]. In their system, dialogue is limited to mobility issues (navigation, obstacle avoidance, etc) with an emphasis on query-response speech acts. They note that a significant side-effect of this type of dialogue is that it can affect how users perceive and treat the robot.

2.2.4 Personality

What is personality?

In psychological terms, personality is the set of distinctive qualities that, taken collectively, distinguishes an individual from another. The majority of research in personality psychology examines the perception of personality, i.e., how other individuals, or the group, perceive a particular individual. The conventional approach is to focus on trait ratings.

Since the late 1980's, the most widely accepted taxonomy of basic individual personality traits has been the "Big Five Inventory" [112]. This taxonomy was developed through lexical analysis of words that are commonly used to describe individual differences. With the "Big Five", an individual's personality can be evaluated in terms of five primary traits:

- extroversion (sociable, outgoing, confidence)
- agreeableness (friendliness, nice, pleasant)
- conscientiousness (helpful, hard-working)
- neuroticism (emotional stability, adjusted)
- openness (intelligent, imaginative, flexibility)

Alternatives to the "Big Five", and other systems based on lexical analysis, include questionnaire-based scales such as the Myers-Briggs Type Indicator (MBTI) and the California Personality Indicator (CPI).

Personality in social robots

There is reason to believe that if a robot had a compelling personality, people would be more willing to interact with it and to establish a relationship with it[27,116]. In particular, personality may provide a useful affordance, giving users a way to model and understand robot behavior[207].

In designing robot personality, there are numerous questions that need to be addressed. Should the robot have a designed or learned personality? Should it mimic a specific human personality, exhibiting (or evoking) specific human traits? Is it beneficial to encourage a specific type of interaction (e.g., infant-caregiver)? What type of personality will best serve a robot's function and capabilities?

There are five common personality types used in social robots:

Tool-like. Used for robots that operate as "smart appliances". Because these robots perform service tasks, they exhibit traits commonly associated with tools (dependability, reliability, etc).

Pet or creature. These toy and entertainment robots exhibit characteristics commonly associated with domesticated animals (cats, dogs, etc.).

Cartoon. These robots exhibit caricatured personalities, such as seen in animation. Exaggerated traits (shyness, stubbornness, etc.) are easy to portray and can be useful for attracting interest and for facilitating interaction with non-specialists.

Artificial being. Inspired by literature and film, primarily science fiction, these robots tend to display artificial (e.g., mechanistic) characteristics.

Human-like. Robots are often designed to exhibit human personality traits. The extent to which a robot must have (or appear to have) human personality depends on its use.

Robot personality is conveyed in numerous ways. Emotions are often used to portray specific stereotype personalities: timid, friendly, etc.[245]. A robot's physical embodiment (size, shape, color), its motion (smooth, jerky, etc.), and how it communicates (sound vs. natural language) also contribute strongly[207]. Finally, the tasks a robot performs may also influence the way its personality is perceived.

2.2.5 Human-oriented perception

To interact meaningfully with humans, social robots must be able to perceive the world as humans do, i.e., sensing and interpreting the same phenomena that humans observe. This means that, in addition to the perception required for conventional functions (localization, navigation, obstacle avoidance), social robots must also possess perceptual abilities similar to humans.

In particular, social robots need perception that is human-oriented: optimized for interacting with humans and on a human level. Social robots need to be able to track human features (bodies, faces, hands). They also need to be capable of interpreting human speech including affective speech, discrete commands, and natural language. Finally, they often must have the capacity to recognize facial expressions, gestures (e.g., hand pointing), and human activity.

Similarity of perception requires more than similarity of sensors (i.e., sensors with characteristics and performance that match human senses). It also is important that humans and robots find the same types of stimuli salient[33]. Moreover, robot perception may need to mimic the way human perception works. For example, the human ocular-motor system is based on foveate vision, uses saccadic eye movements, and exhibits specific visual behaviors (staring vs. glancing, breaking contact, etc.). Thus, in order for a robot to be readily understood, its visual motor control may need to have similar characteristics[31,35,27].

Human-oriented perception is an important research topic for applications other than social robots. Human tracking is considered to be an essential component of intelligent environments and smart spaces (e.g., [230]). Speech and gesture recognition is an integral component of multimodal and perceptual user interfaces [118]. Activity detection and recognition plays a fundamental role in automated surveillance[174,49].

Types of perception

Most human-oriented perception is based on passive sensing, typically computer vision and spoken language recognition. Passive sensors, such as CCD cameras, are cheap, require minimal infrastructure, and can be used for a wide range of perception tasks[82], [54], [2], and [95].

Active sensors (ladar, ultrasonic sonar, etc.), though perhaps less flexible than their passive counterparts, have also received attention. In particular, active sensors are often used for detecting and localizing human in dynamic settings.

People tracking

For human-robot interaction, the challenge is to find efficient methods for people tracking in the presence of occlusions, variable illumination, moving cameras, and varying background. A broad survey of human tracking is presented in [95]. Specific robotics applications can be reviewed in [223], [164], [37], and [186].

Speech recognition

Speech recognition is generally a two-step process: signal processing (to transform an audio signal into feature vectors) followed by graph search (to match utterances to a vocabulary). Most current systems use Hidden Markov Models to stochastically determine the most probable match. An excellent introduction to speech recognition is [190].

Human speech contains three types of information: who the speaker is, what the speaker said, and how the speaker said it[27]. Depending on what information the robot requires, it may need to perform speaker tracking, dialogue management, or emotion analysis. Recent applications of speech in robotics include [27], [133], [146], [173], and [216].

Gesture recognition

When humans converse, we use gestures to clarify speech and to compactly convey geometric information (location, direction, etc.). Very often, a speaker will use hand movement (speed and range of motion) to indicate urgency and will point to disambiguate spoken directions (e.g., "I parked the car over there").

Although there are many ways to recognize gestures, vision-based recognition has several advantages over other methods. Vision does not require the user to master or wear special hardware. Additionally, vision is passive and can have a large workspace. Recognizing gestures is a complex task that involves motion modeling, motion analysis, pattern recognition, and machine learning. Two excellent overviews of vision-based gesture recognition are [181] and [243]. Details of specific systems appear in [126], [244], and [236].

Facial perception

Face detection and recognition. A widely used approach for identifying people is face detection. Two comprehensive surveys are [48] and [88]. A large number of real-time face detection and tracking systems have been developed in recent years, such as [229] and [202,203].

Facial expression. Since Darwin[55], facial expressions have been considered to convey emotion. More recently, facial expressions have also been thought to function as social signals of intent. A comprehensive review of facial expression recognition (including a review of ethical and psychological concerns) is [136]. A survey of older techniques is [199].

There are three basic approaches to facial expression recognition [136]. Image motion techniques identify facial muscle actions in image sequences. One problem with this approach is that head motion involves both rigid and non-rigid motion, which can be difficult to separate. Anatomical models track facial features, such as the distance between eyes and nose. A limitation of this approach is that it is hard to construct a model that remains valid across many individuals. Principal component analysis (PCA) reduce image-based representations of faces into principal components such as eigenfaces or holons. The primary difficulty with these appearance-based approaches is that they are very sensitive to imaging issues including illumination, head orientation, and sensor noise.

Gaze tracking. Gaze is a good indicator of what a person is looking at and paying attention to. A person's gaze direction is determined by two factors: head orientation and eye orientation. Although numerous vision systems have been developed to track head orientation (generally based on frontal faces), few researchers have attempted to track eye gaze using only passive vision. Furthermore, such trackers have not proven to be dependable nor highly accurate[229]. Gaze tracking research includes [202] and [221].

2.2.6 User modeling

In order to interact with people in a human-like manner, socially interactive robots must perceive and understand the richness and complexity of natural human social behavior[27]. Detecting and recognizing human action and communication provides a good starting point. More important, however, is being able to interpret and reacting to human behavior. A key mechanism for performing this is user modeling.

User modeling can be quantitative, based on the evaluation of parameters or metrics. The stereotype approach, for example, classifies users into different subgroups (stereotypes), based on the measurement of pre-defined features for each subgroup[225]. User modeling may also be qualitative in nature. Interactional structure analysis, story and script based matching, and BDI all focus on identifying subjective aspects of user behavior.

There are many types of user models: cognitive, emotional, attentional, etc. A user model generally contains a set of attributes that describe a user, or group, of users. Models may be static (defined *a priori*) or dynamic (adapted or learned). Information about users may be acquired explicitly (through questioning) or implicitly (inferring through observation). The former can be time consuming and the latter can be difficult, especially if the user population is diverse or has broad characteristics[99].

User models are employed for a variety of purposes. First, user models help the robot understand human behavior and dialogue. Many spoken dialogue systems depend on user models for recognizing and generating speech acts. Second, user models shape and control feedback given to the user. Robot motions (body position, gaze direction, etc.) and interaction pacing (e.g., knowing when to insert pauses) can be directed by appropriate models. Finally, user models are useful for adapting the robot's behavior to accomodate users with varying skills, experience, and knowledge.

Fong, Thorpe and Baur employ a stereotype user model to adapt human-robot dialogue and to modify robot behavior to fit the needs of different users[84]. This occurs in three ways: the questions the robot asks to the human are based on an estimate of the user's knowledge; the weight the robot gives to human responses are based on an estimate of the user's ability to provide accurate responses; and the independence the robot exhibits (i.e., its level of autonomy) is adapted to fit human decision making (e.g., how quickly a user is able to respond to safety critical questions).

Pineau et al. discuss social interaction between a nurse robot and elderly individuals[186]. Because many elderly have difficulty understanding synthesized speech, as well as articulating responses in a computer understandable way, the robot needs to adapt to individuals. To do this, a cognitive orthotic system, "Autominder", is used to provide individual-specific reminders about daily activities. Autominder maintains a model of the user's daily schedule, monitors performance of activities, and plans reminders accordingly. The user model is a Quantitative Temporal Bayes Net, which can reason about probabilistic temporal constraints.

Schulte et al. describe adaptation in Minerva, a tour guide robot[206]. Minerva employs a memory-based learner for learning how to interact with different people. In particular, between museum tours, Minerva enters an "attraction interaction" state, in which the goal is to attract people in preparation for the next tour. Using a continuous adaptation strategy, Minerva autonomously learned to select actions (speech acts, facial expressions, etc.) that improved its ability to attract user interest.

2.2.7 Socially situated learning

In socially situated learning, an individual interacts with his social environment to acquire new competencies. Humans and some animals (e.g., primates) learn through a variety of techniques including direct tutelage, observational conditioning, goal emulation, and imitation[93]. One prevalent form of influence is local, or stimulus, enhancement in which a teacher actively manipulates the perceived environment to direct the learner's attention to relevant stimuli[139].

Robot social learning

For social robots, learning is used for transferring skills, tasks, and information. Learning is important because the knowledge of the teacher, or model, and robot may be very different. Additionally, because of differences in sensing and perception, the model and robot may have very different views of the world. Thus, learning is often essential for improving communication, facilitating interaction, and sharing knowledge [119].

A number of studies in robot social learning have focused on robot-robot interaction. Some of the earliest work focused on cooperative, or group, behavior[143,9]. A large research community continues to investigate group social learning, often referred to as "swarm intelligence" and "collective robotics". Other robot-robot work has addressed the use of "leader following" [103,56], inter-personal communication[18,16,217], imitation[17,94], and multi-robot formations[156].

In recent years, there has been significant effort to understand how social learning can occur through human-robot interaction. One approach is to create sequences of known behaviors to match a human model [145]. Another approach is to match observations (e.g., motion sequences) to known behaviors, such as motor primitives [73,74]. Recently, Kaplan et al. have explored the use of animal training techniques for teach an autonomous pet robot to perform complex tasks[113]. The most common social learning method, however, is imitation.

Imitation

Imitation is an important mechanism for learning behaviors socially in primates and other animal species[67]. At present, there is no commonly accepted definition of “imitation” in the animal and human psychology literature. An extensive discussion is given in [102]. Researchers often refer to Thorpe’s definition[227], which defines imitation as the “copying of a novel or otherwise improbable act or utterance, or some act for which there is clearly no instinctive tendency”.

With robots, imitation relies upon the robot having many perceptual, cognitive, and motor capabilities[34]. Researchers often simplify the environment or situation to make the problem tractable. For example, active markers or constrained perception (e.g., white objects on a black background) may be employed to make tracking of the model amenable.

Breazeal and Scassellati argue that even if a robot has the skills necessary for imitation, there are still several questions that must be addressed[34]:

- *How does the robot know when to imitate?* In order for imitation to be useful, the robot must decide not only when to start/stop imitating, but also when it is appropriate (based on the social context, the availability of a good model, etc).
- *How does the robot know what to imitate?* Faced with a stream of sensory data, the robot must decide what actions are worth imitating. Furthermore, the robot must determine which of the model’s actions are relevant to the task, which are part of the instruction process, and which are circumstantial.
- *How does the robot map observed action into behavior?* Once the robot has identified and observed salient features of the model’s actions, it must ascertain how to reproduce these actions through its behavior.
- *How does the robot evaluate its behavior, correct errors, and recognize when it has achieved its goal?* In order for the robot to improve its performance, it must be able to measure to what degree its imitation is accurate and to recognize when there are errors.

Imitation has been used as a mechanism for learning simple motor skills from observation, such as block stacking[131] or pendulum balancing[204]. For these tasks, the robot must be capable of observing and replicating geometric relationships (e.g., relative arm position). One way to do achieve this is to use movement matching, such as discussed by Demiris and Hayes[73].

Andry et al. have used imitation to speed up the learning of sensor-motor associations[4]. In their system, a neural network architecture enables a robot to perform low-level imitations (i.e., reproducing simple and meaningless movements). Imitative behavior is triggered by perception ambiguity and enables the system to learn new sensor-motor associations without explicit reinforcement.

Nicolescu and Mataric describe a robot that follows a human teacher, gathering observations from which it constructs task representations (a network of robot behaviors)[168]. The ability to learn from observation is based on the robot’s ability to relate the observed states of the environment to the known effects of its own behaviors.

2.2.8 Intentionality

Dennett contends that humans use three strategies to understand and predict behavior[72]. The *physical stance* (predictions based on physical characteristics) and *design stance* (predictions based on the design and functionality of artifacts) are sufficient to explain simple devices. With complex systems (e.g., humans), however, we often do not have sufficient information, to perform physical or design analysis. Instead, we tend to adopt an *intentional stance* and assume that the systems’ actions result from its beliefs and desires.

In order for a robot to interact socially, therefore, it needs to provide evidence that is intentional (even if it is not intrinsically [201]). For example, a robot could demonstrate goal-directed behaviors, or it could exhibit the attentional capacity. If it does so, then the human will consider the robot to act in a rational manner.

Attention

Scassellati discusses the recognition and production of joint attention behaviors in Cog[202]. Just as humans use a variety of physical social cues to indicate which object is currently under consideration, Cog performs gaze following (detecting eye contact and extracting angle of gaze), imperative pointing (foveating a target and ballistic reach), and declarative pointing (simple mimicry of gestures such as head nods).

Kopp and Gardenfors also claim that attentional capacity is a fundamental requirement for intentionality[125]. In their model, a robot must be able to identify relevant objects in the scene, direct its sensors towards an object, and maintain its focus on the selected object.

Marom and Hayes (1999, 2001a,b) consider attention to be a collection of mechanisms that determine the significance of stimuli[139–141]. Their research focuses on the development of pre-learning attentional mechanisms, which help reduce the amount of information that an individual has to deal with.

Expression

Kozima and Yano argue that to be intentional, a robot must exhibit goal-directed behavior[122,123]. To do so, it must possess a sensorimotor system, innate reflexes (behaviors), drives that trigger these behaviors, a value system for evaluating perceptual input, and an adaptation mechanism.

Breazeal and Scassellati describe how Kismet conveys intentionality through motor actions and facial expressions[32]. In particular, by exhibiting proto-social responses (affective, exploratory, protective, and regulatory) in the same manner as infants, the robot provides cues that enable adults to interpret its actions as intentional.

Schulte et al. discuss how a caricatured human-face and simple emotion expression can convey intention during spontaneous short-term interaction[206]. For example, a tour guide robot might have the intention of making progress while giving a tour. Its facial expression and recorded speech playback can communicate this information.

3 Applications

3.1 Robots as test subjects

Evaluating models of social and biological development is difficult in natural settings. Ethical concerns, complications preventing test procedure implementation, and difficulties isolating hypothesized variables often make experimental evidence difficult (or impossible) to obtain. As a result, numerous researchers have begun exploring how social robots can serve as experimental test subjects[202,203].

In particular, robots are increasingly being used as research tools to perform a wide range of comparative studies[64,4]. To date, robots have been used to examine, validate and refine theories of social and biological development, psychology, neurobiology, emotional and non-verbal communication, and social interaction.

There are are numerous reasons why using robots as test subjects is useful, especially compared with computational simulation[202,218]:

- Implementing a model for physical experimentation requires specifying, in detail, all internal structures and processes. This makes it explicitly clear how all theoretical assumptions have been operationalized.
- Internal model structures can be manipulated and examined. This is not possible with human beings, except via indirect measures (e.g., subject questioning).
- Experiments can be repeated with nearly identical conditions and small variations can be used to isolate single factors. This is difficult to achieve with conventional, observation-based studies.

- We can easily examine alternative hypotheses. For example, we can compare what an individualistic inductive learning process would achieve with the same data as a social learning process.
- Robots can be subjected to testing that may be hazardous, costly, or unethical to conduct on humans. This includes assessing intervention strategies or treatments, which may be controversial to test on human subjects.

At the same time, there are obviously also important limits to this methodology. For example, current robots cannot fully represent human (even infant) cognition, perception and behavior. Thus, for now, the most that experimentation on robots can achieve is examination of specific (or limited) model features.

3.1.1 Social development

Perhaps the most obvious use of social robots as test subjects is to examine theories of social development. For the most part, the focus of these experiments is to confirm, or to refute, claims of how infants and young children develop in social learning skills such as imitation and joint attention.

Infanoid is an infant-like humanoid robot (see Figure 34) used to investigating the underlying mechanisms of social development [121,122]. In particular, it is being used to explore how socially meaningful behaviors develop through physical and socio-cultural interaction with human caregivers. Infanoid is designed to have the same size and kinematic structure of a three-year-old child's upper body.

Cog is a general-purpose humanoid platform (see Figure 36) intended for exploring theories and models of intelligent behavior and learning[1]. Cog's design implements two popular theories for the development of theory of mind. Scassellati describes experimentation with a partial implementation of Baron-Cohen's model of social development[10] that focuses on the recognition and production of joint attention behaviors[202,203].

Andry et al. present an imitative robot architecture that is designed to exhibit different phases of social development, comparable to those seen in children[4]. They are using this architecture to test psychological and neurbiological models in an effort to understand mental development problems (e.g., autism) as well as to ascertain the fundamental properties of imitation games engaged in by children.

3.1.2 Social interaction

Another area of active experimentation with robots is social interaction. In these studies, the primary emphasis is to examine theories that describe how interaction in a social context influences cognition. Of secondary interest is exploring how social interaction provides a basis for symbol grounding and communication.

Since its creation in 1997, Kismet has been used to examine various aspects of infant-caregiver interaction[22,27]. In numerous studies, a wide range of infant-level social competencies have been examined, including turn-taking, low-level emotional exchanges, and acquisition of meaningful communication acts.

Dautenhahn and Billard discuss experiments carried out in teaching a synthetic language-like system to a robot[64]. These experiments were designed to test claims made in Vygotsky's theory of child cognitive development, i.e., that social interactions are fundamental to initial cognitive development and provide a context that can scaffold richer cognitive functions.

In order to identify the key factors in non-verbal communication, Aoki et al. investigated interactions between laboratory rats and rat-like robots[5]. They report that the movement of the robot acts as stimulation for discriminative learning, enabling it to guide real rats in forming associations such as social pecking order. Aoki et al. contend that studies of rat-like robots will lead to better understanding of human behavior, in the same manner that animal experimentation (especially on rats) does.

3.1.3 Emotion, attachment and personality

Social robots have also been used to examine and validate a variety of theories of individual behavior. Nakata et al. conducted a study of robot movement using theories of dance psychology[163]. In this study, they attempted to verify predictions of how various types of movement are perceived by humans working in close proximity. A primary objective was to better understand how observed body movement may affect the emotional state of human.

Shibata et al. have begun investigating how emotional behaviors develop through long-term physical contact and interaction[209–211]. They contend that when a human and a pet robot interact, they mutually stimulate and affect each other. Furthermore, they claim that as consequence of this interaction, that humans can develop the same attachment to the robot as to a real pet, especially if the relationship occurs over the long-term.

Miwa et al. implemented a number of robot personalities on a head robot[159]. The personality model was directly based on one of the "Big Five" taxonomies of basic individual personality traits. They then conducted a series of tests to verify how different personalities can influence, and be influenced by, various emotional and perception models.

3.2 Service

Robots are clearly of great functional value when they provide concrete services for humans. For example, welding robots have significantly increased throughput in automotive assembly plants by providing very fast and precise welding services for the assembly line. In this section, we examine research projects in which a social robot is designed and tested for a task-focused mission.

There are various reasons why a task-oriented robot can find social interaction to be beneficial. Perhaps the most evident is usability. In any case where a robot is to interact directly with one or more humans, social engagement such as spoken dialogue and gestures may help to make the robot easier-to-use for the novice and also more efficient-to-use for the expert. One psychological mechanism that also falls within this class involves the comfort level of humans interacting with the robot. Social interaction can be designed to make the humans more comfortable as they share a space with the robot, just as appropriate social interaction does so between humans sharing a space.

Fig. 15. Pearl is a Nursebot, providing cognitive and spatial aide

This section proceeds by describing several research projects that inhabit three different categories. First, we examine the robot whose role is to aid one or more humans over the long term. Second, we turn to robots that serve as members of a larger team with a common goal. Finally, we examine robots whose tasks require spontaneous, short-term interaction with large numbers of humans in public spaces.

3.2.1 Robots as durative assistants

One important reason for Japan's large investment in personal robotics has to do with the significant proportion of elderly Japanese, both today and increasingly in future years. Robots that serve as aides to the elderly may prolong the autonomy of individuals, staving off a move to managed care facilities for months or even years.

Pineau et al. describe a robotic assistant for nursing homes that is just such a durative aide to the elderly[186]. The robot, Pearl, has two main functions: to serve as a cognitive orthotic to its users; and to serve as a spatial guide for navigating the environment (Figure 15). As a cognitive orthotic, Pearl can remind individuals to use the restroom, eat and turn on the television for a favorite show. As a guide, Pearl can not only remind one of a physical therapy appointment, but can also lead the way to the exercise room, for instance.

The job of Pearl is heavily social, and so basic skills required by the robot include social perception (i.e. people-tracking, speech recognition) as well as social closed-loop control (i.e. dialogue management, leading the way while matching walking speed, reminding and verifying). Dialogue management includes reasoning and acting based on knowledge or lack thereof. For instance Pearl is able to ask clarification questions if it failed to understand a persons directive or response. The authors are able to show that such adaptive dialogue improves overall error rates for the robot considerably.

Severinson-Eklundh et al. have also implemented and tested a robot that can serve as a long-term aide to one or more individuals[207]. In this case, rather than serving a cognitive function, the robot is meant to provide fetch-and-carry services for motion-impaired users in the office environment. As a social agent, this robot will not only have a social relationship with the user but also with all other persons in the office building, who will encounter this robot as it performs its activities.

Following their ethic of task-based and context-based design, Severinson-Eklundh et al. identified two critical communicative needs of a fetch-and-carry robot. First, the robot must be able to signal its comprehension of commands given to it. Second, the robot must be able to indicate the direction in which it desires to navigate. In particular, given the round Nomad Scout platform, the question of direction of motion proved to be important for bystanders who wish to help the robot by getting out of its way.

Fig. 16. eMuu (left) and Muu2 (right)

Figure 10 shows the CERO representative mounted on the mobile platform. By creating a social representative for the whole robot, Severinson-Eklundh takes a fascinating approach to social robot design, in which a focal point, in this case CERO, represents the social interactive aspects of the whole.

Bartneck suggests similarly that robotic social agents may serve as representatives between humans and complex systems[11,12]. In this case, eMuu and Muu2 are proposed as interfaces between a homeowner and the smart appliances in the home, or even the intelligent home itself. As with CERO, the robots shown in Figure 16 demonstrate a departure from anthropomorphic design while still achieving engagement at the social level.

3.2.2 Robots as collaborators

As another application groups of humans and robots will someday together comprise teams with a common goal. Thus the robot becomes, not just an assistant, but a partner in accomplishing the team objectives. Fong et al. propose that such robots will maximize their contributions by being able to take advantage of human skills and human advice and decision-making as appropriate[84]. Thus although there is no reason for such robots to be viewed as peers to humans, they must nonetheless be active partners, recognizing their limitations and requesting help as needed.

Fong proposes four key qualities that a collaborator robot must have in order to take part in such a social niche. First, the robot must have sufficient self-awareness, or introspection, to recognize its limitations. This is critical in making the determination to request human help. Second, the robot must be self-reliant at the most basic level. Thus it should be able, in any condition, to safe itself and thus

Fig. 17. Collaborative control. Robot in need of assistance (left) and its query to the human (right).

avoid the hazards that it may face. Third, the robot must have dialogue competence so that two-way information exchange between robot and human is possible. Finally the robot should be adaptive, and thereby able to make use of all variety of human resources, from the robot-novice to the experienced human team members.

The coupling of human to robot in this collaboration model is many-to-many. Groups of humans and groups of robots may be part of the system, and it is through their collaboration with one-another that the group goals are achieved. This model breaks out of the master-slave control formula, due to several important limitations. First, the master-slave control architecture fixes the robots level of autonomy and thereby the degree of control exerted by the human master. Furthermore, such an architecture tends to enforce a fixed relationship, not only through time between a single master and slave, but also as implemented for various masters.

As an alternative to master-slave control, Fong et al. suggests collaborative control, a system model based on human-robot dialogue[84]. With this model, the robot asks questions to the human in order to obtain assistance with cognition and perception (Figure 17). This enables the human to function as a resource for the robot and to help compensate for limitations of autonomy.

In [85], Fong et al. describe an experiment in which robot-human dialogue is used to achieve remote robotic exploration, a task traditionally performed with direct teleoperation. Using a contextual inquiry based approach, the authors present both expert and novice robot-users with a collaborative control scenario in which a room not directly visible must be explored by a robot-human team. The use of simulated error conditions (e.g. temperature sensor fluctuations) enriches the dialogue between robot and human, as the robot requests clarification of its safe operational parameters during execution.

3.2.3 Short-term public interaction robots

In contrast to both personal assistant robots and collaborative robots, an important quality of public interaction robots is that the human-robot relationship is uninitiated, untrained and very short-term. Indeed, by public we also imply that the variance among potential human targets is quite high: some will be more technologically savvy and robot-friendly than others. Finally, unlike the previous categories public spaces are not well-suited to gating social interaction for one-on-one relationships. Crowds naturally form, and interested persons cluster automatically. Thus, one-on-one engagement alone is unacceptable.

The task-based robots that we consider in this section all have missions revolving about information: either the deployment of information to the public, or the acquisition of information about the public. Thus the challenge in the former case is to deliver the information effectively to the right persons; and in the latter case to collect the desired information from the right persons.

A number of social robots have been designed with specific public missions at hand. These robots have all demonstrated that, with the right behavior, mobile robots can successfully attract, engage and interact with the public in busy spaces.

A recent application of public robotics is the Robox series of robots (Figure 18) installed at Switzerlands Expo 2002[213]. During this national exposition, the public robots were charged with two goals: present themselves as demonstrations of robotic

Fig. 18. The Robox installation at Switzerlands Expo 2002

technology; and guide visitors through a roomful of robotic exhibits, presenting these external robotic exhibits on the nature of robotic technology.

The physical design of the exhibition robot reflects its social charter. Using eyebrows and asymmetric eyes, each robot is capable of affective expression as well as communication of intentionality- such as its desired direction of travel for example. Sensors for human detection and tracking include laser rangefinding, color vision and tactile sensors.

At the heart of each robot's behavior are interaction sequences which lead humans through the exhibit space and provide tour content. To implement this sequential tour controller, while also performing adaptation based on the behavior of the humans around each robot, a hybrid controller sequences static interaction sequences while monitoring sensor conditions that would trigger asynchronous departures from the nominal interaction. For instance if a visitor blocks the robots path repeatedly, then the static sequence is interrupted by a dynamic interaction sequence designed to convince the visitors to allow the robot to do its job.

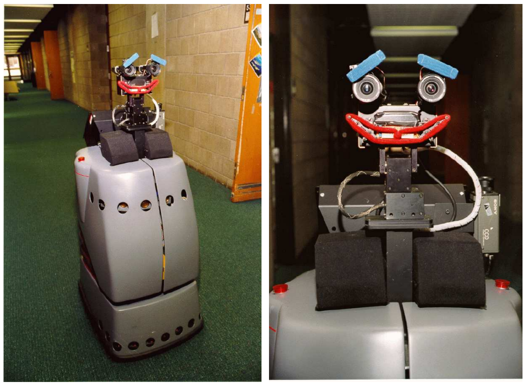

Fig. 19. Minerva (Carnegie Mellon University)

In earlier work other researchers have demonstrated that affect can have a measurable, quantitative influence on the efficacy of the robot at providing tours to the public. Schulte et al. describe the critical differences between the Minerva tour guide robot (Figure 19), which provided tours for a two-week period at the Smithsonian Institutions Museum of American History, and an earlier robot, Rhino, which provided tours in a German science museum[206].

In [206], three characteristics are suggested as critical to the success of robots that must exhibit spontaneous interactions in public settings. First, the robot must include a focal point, which serves as an obvious focus of attention for the human. A robotic face or animated face serves this function on many tour-guide robots. Second, the robot should communicate emotional state to the visitors as a way of efficiently conveying its intention- for example its frustration at being blocked by playful tourists. Third, the robot should have the capability to adapt its human interaction parameters based on the outcome of past interactions so that it can continue to demonstrate open-ended behavior as the visitors in its environment change.

Minerva uses a motorized facial caricature consisting of eyes, lips and eyebrows to provide both a focal point and a means for the communication of emotional state. A simple state transition diagram modulates facial expression and speech, from happy to neutral to angry, based on the length of time for which Minervas path has been blocked. The authors notes that Rhino, a robot with no such face but with relatively similar navigation means, achieved average tour speeds approximately 20% slower than Minerva, thus providing some empirical evidence as to the efficacy of affective interaction.

Fig. 20. The Mobot robots Chips (top left), Sweetlips (right) and Joe Historybot (bottom left)

The Mobot Museum Robot series of robots, comprised of four robots deployed and tested sequentially over a five-year period, forms a basis for observations about the evolution of public, social robots[241]. The goals of these robot installations were unchanging: deployment of mobile robots in museum spaces that bring about compelling and fruitful interactions between the robots, human visitors and the exhibits in the space. Yet, with the re-design of subsequent robots significant trends were identified by the authors pertaining to lessons learned with each deployment.

In terms of interaction content, the authors studied a gradual evolution from one-way, didactic teaching to short-term, two-way, challenge-based interaction (Figure 20). The first robot, Chips (also known as "Sage"), provided two- and three-minute long tour speeches at special locations along its travel path. The user faced a single button and was only capable of asking the robot to present or skip each such tour stop.

The second robot, Sweetlips, equipped with five buttons, encouraged the visitor to select from several tour themes (e.g. predator-prey relationships; animals and their young; etc.). Depending on the visitors choice, a somewhat customized tour with only 30-second tour stops would proceed.

The third robot, Joe Historybot, sported a touch screen, offering not only customized tours, but also secondary information about the space (e.g. location of other exhibits and the restrooms). Furthermore, this robot provided, not unidirectional tours but rather a series of puzzles and questions along the tour paths, encouraging the visitors to stay engaged and challenged throughout the tour process.

The primary lesson learned from the Mobot series is that user learning is more effective and enjoyable when there is strong interaction and dialogue between robot and human. In particular, information is more effectively conveyed to humans when given within the context of two-way questions and answers.

A challenge in assessing robot-human interaction in public spaces revolves is the lack of appropriate controls. Each public installation occupies a new space with different cross-sections of society. Thus, it is difficult to quantitatively factor context out of the equation in order to arrive at well-justified conclusions regarding the social behavior appropriate for a mobile robot.

In work aimed at overcoming this issue, Bruce et al. describe factored experiments in which the same social robot, Vikia (Figure 38), is placed in the same public environment repeatedly[37]. Each experimental instance inactivates some portion of Vikias social competencies (e.g. animated visual saccades; physical head panning, facial animation; speech) and measures the robots social efficacy, in this case its penetration rate at engaging passers-by for a brief, anonymous poll. This work demonstrates the practicality of using formal, psychological techniques to explore the interaction mechanisms at work when robot are tasked with establishing effective, short-term relationships with the public.

Future work in the various contexts of task-oriented robotics, from personal assistants and collaborators to public interaction systems, will doubtless lead to richer and more meaningful social relationships between robots and humans for years to come.

3.3 Entertainment, personal, and toys

Commercially viable robotic pursuits in the toy and entertainment markets are governed by several constraints. An entertaining robot must achieve a maximum of entertainment value at a minimum of cost, which means that sensors must be multi-purpose and effectors must be minimized. Thus it is generally true that toy robots are technologically subpar as compared to industrial and research robot platforms, and furthermore the technology curve for toy robots will tend to be shallower.

But the toy market is most adept at applying play-pattern based design principles to the problem of toy-human interaction. For a given toy, designers identify a finite collection of ways in which the user is intended to interact with the toy. The toy is then designed with the explicit goal of enabling the desired list of play patterns.

3.3.1 Animatronic children's dolls

The children's doll continues to be one of the best-selling toys in history, and as such, toy manufacturers continue to explore new avenues for distinguishing their product from those of their competition. One of many heuristics states that a high-volume toy cannot be priced more than $99 retail, which implies a manufacturing cost of approximately $20. With the cost of low-end microprocessors, MEMS-based accelerometers, voice recognition chips and imaging chips continually decreasing, it has become possible to develop robotic dolls.

It was with these price trends in mind that Hasbro funded IS-Robotics (now known as iRobot) to design an animatronic children's doll that provides palpable interactivity in place of the purely "pretend" interactivity of a conventional doll.

My Real Baby is the toy resulting from this endeavor (Figure 21). The development process focused most closely on the doll's face, using small actuators planted underneath the rubber face to distort the face along several degrees of freedom, producing smiling, laughing and grimacing expressions. In addition to expression, the robot uses a