

58. Social Robots that Interact with People

Cynthia Breazeal, Atsuo Takanishi, Tetsunori Kobayashi

This chapter surveys some of the principal research trends in social robotics and its application to human–robot interaction (HRI). Social (or sociable) robots are designed to interact with people in a natural, interpersonal manner – often to achieve social–emotional goals in diverse applications such as education, health, quality of life, entertainment, communication, and collaboration. The long-term goal of creating social robots that are competent and capable partners for people is quite a challenging task. They will need to be able to communicate naturally with people using both verbal and nonverbal signals. They will need to engage us not only on a cognitive level, but on an emotional level as well. They will need a wide range of social–cognitive skills and a theory of other minds to understand human behavior, and to be intuitively understood by people. A deep understanding of human intelligence and behavior across multiple dimensions (i. e., cognitive, affective, physical, social, etc.) is necessary in order to design robots that can successfully play a beneficial role in the daily lives of people. This requires a multidisciplinary approach where the design of social robot technologies and methodologies are informed by robotics, artificial intelligence,

58.1 Social Robot Embodiment.....	1350
58.2 Multimodal Communication	1352
58.2.1 Robots that Express Paralinguistic Information.....	1352
58.2.2 Robots that Understand Paralinguistic Information.....	1353
58.2.3 Group Conversation	1355
58.2.4 Communication in Collaboration ...	1355
58.3 Expressive Emotion–Based Interaction....	1356
58.3.1 Kismet: Inspiration from Developmental Psychology ...	1356
58.3.2 WE-4RII: A Model of Emotion Inspired by Motion	1357
58.4 Socio–cognitive Skills	1360
58.4.1 Shared Attention.....	1360
58.4.2 Emotional Empathy	1362
58.4.3 Mental Perspective Taking	1362
58.4.4 Perspective Taking in Collaboration	1363
58.5 Conclusion and Further Reading	1365
References	1366

chology, neuroscience, human factors, design, anthropology, and more.

The way a person interacts with a social robot (or sociable robot) is quite different from interacting with the majority of autonomous mobile robots today. Modern autonomous robots are generally viewed as tools that human specialists use to perform hazardous tasks in remote environments (i. e., sweeping minefields, inspecting oil wells, mapping mines, etc.). In dramatic contrast, social (or sociable) robots are designed to engage people in an interpersonal manner, often as partners, in order to achieve social or emotional goals.

The development of socially intelligent and socially skillful robots drives research to develop autonomous or semiautonomous robots that are natural and intuitive for the general public to interact with, communicate with, work with as partners, and teach new capabilities. *Dautenhahn's* work is among the earliest in thinking about robots with interpersonal social intelligence where relationships between specific individuals are important [58.1, 2]. These early works pose the question *What are the common social mechanisms of communication and understanding that can produce efficient, enjoyable,*

natural, and meaningful interactions between humans and robots?

Promisingly, there have been initial and ongoing strides in all of these areas ([58.3–11], etc.). In addition, this domain motivates new questions for robotics researchers, such as how to design for a successful long-term relationship where the robot remains appealing and provides consistent benefit to people over weeks, months, and even years. The benefit that social robots provide people extends far beyond strict task performing utility to include educational (Chap. 55), health and therapeutic (Chap. 53), domestic (Chap. 54), social and emotional goals (e.g., entertainment, companionship, communication, etc.), and more.

58.1 Social Robot Embodiment

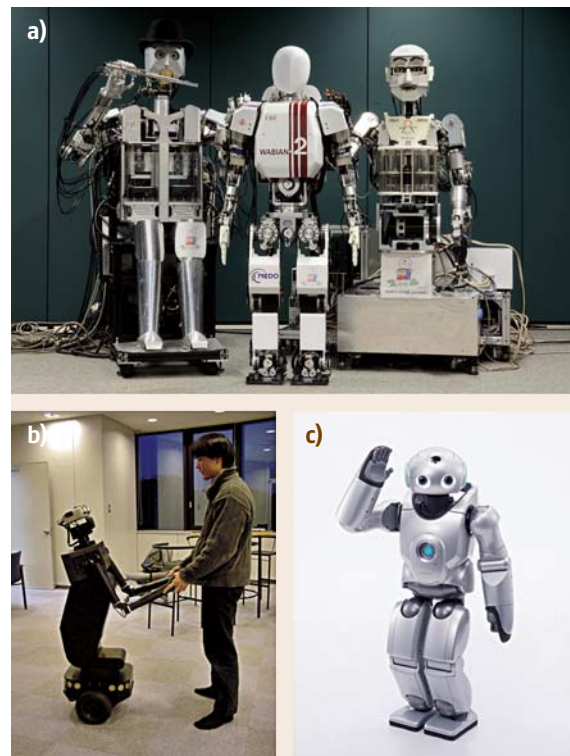
Social robots are designed to interact with people in human-centric terms and to operate in human environments alongside people. Many social robots are humanoid or animal-like in form, although this does not have to be the case. A unifying characteristic is that social robots engage people in an interpersonal manner, communicating and coordinating their behavior with humans through verbal, nonverbal, or affective modalities. As can be seen in the following examples, social robots exploit many different modalities to communicate and express social-emotional behavior. These include whole-body motion, proxemics (i.e., interpersonal distance), gestures, facial expressions, gaze behavior, head orientation, linguistic or emotive vocalization, touch-based communication, and an assortment of display technologies.

For social robots to close the communication loop and coordinate their behavior with people, they must also be able to perceive, interpret, and respond appropriately to verbal and nonverbal cues from humans.

Fig. 58.1a–c Examples of socially interactive humanoid robots: **(a)** humanoid robots developed at Waseda University from *left to right*: a flautist robot WF-4RII [58.12], **WABIAN-2** [58.13], and WE-4RII [58.14] **(b)** Robovie III developed by the Advanced Telecommunications Research Institute International (ATR) Intelligent Robotics and Communication Laboratories, which is able to gesture with its arms and give a hug **(c)** quest for curiosity (QRIO), a small biped entertainment robot previously developed by Sony, is very well known for its impressive dancing ability

We begin this chapter with a brief overview of a wide assortment of socially interactive robots that have been developed around the world (Sect. 58.1). We follow with selected topics that highlight some of the representative research themes: multimodal communication (Sect. 58.2), expressive emotion-based interaction (Sect. 58.3), and social-cognitive skills (Sect. 58.4). We rely on examples from our own research programs to illustrate these trends, while making reference to other excellent work performed in other research labs. The following robotic platforms are used as case studies through the sections: Kismet and Leonardo developed at MIT, and Waseda Eye No. 4 Refined II (WE-4RII), ROBISUKE, and ROBITA developed also at Waseda University.

Given the richness of human behavior and the complexity of human environments, many social robots are among the most sophisticated, articulate, behaviorally rich, and intelligent robots today.



As shown in Fig. 58.1, a number of socially interactive humanoid robots have been developed (see Chap. 56) that can participate in whole body social interaction with people such as dancing [58.15], walking hand-in-hand [58.16,17], playing a musical duet [58.12], or transferring skills to unskilled persons [58.18]. Their arms and hands are designed to exhibit human-like gestures such as pointing, shrugging shoulders, shaking hands, or giving a hug [58.19–21]. Some of them are designed with mechanical faces to communicate with humans via facial expressions.

Whereas many of these humanoids have a mechanical appearance, android robots are designed to have a very human-like appearance with skin, teeth, hair, and clothes (Fig. 58.2). A design challenge of android robots is to avoid the *uncanny valley* where the appearance and movement of the robot resemble more of an animate corpse than a living human. Designs that fall within the *uncanny valley* elicit a strong negative reaction from people [58.30].

There are a number of more creature-like social robots that take their aesthetic and behavioral inspiration from animals (Fig. 58.3). Given that people pet and stroke companion animals, touch-based communication has been explored in several of these more animal-inspired robots. Sony's entertainment robot dog, AIBO [58.26], is a well-known commercial example. Other robots in this category have a more organic appearance, such as the therapeutic companion robot seal, Paro [58.27]. Researchers have also chosen to design robots with a more fanciful appearance, melding anthropomorphic with animal-like qualities such as Leonardo ([58.28, 29, 31], etc.).

Many social robots are not overtly humanoid or zoomorphic, but still capture key social attributes (Fig. 58.4), for instance, one of the best-known and pioneering social robots Kismet [58.3] developed at the MIT Artificial Intelligence Lab. Kismet had a very expressive mechanical face with anthropomorphic features like large blue eyes. Another example is the dancing robot Keepon developed by National Institute of Information and Communications Technology (NiCT) (Japan). This small yellow robot has a simplistic face and uses a classic animation technique called *squash and stretch* for expression of the body [58.32].

Many mobile social robots have been fitted with faces to enhance social interaction (Fig. 58.4). Some examples are the elder-care robot, Pearl, and the robotic receptionist Valerie with a graphical face on a liquid-crystal display (LCD) screen [58.33], both developed at Carnegie Mellon University [58.34]. Other examples are

robots like PaPeRo developed by NEC, aiming at a commercial product [58.35]. Still, some social robots have

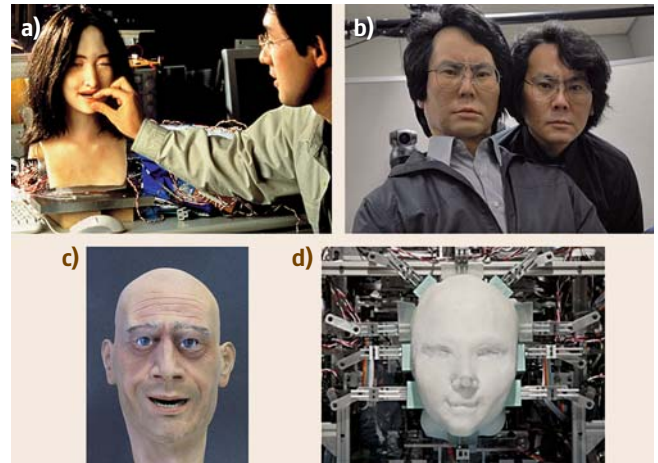


Fig. 58.2a–d Some examples of androids: (a) one of the earliest face robots developed at the Science University of Tokyo [58.22] (b) Geminoid HI-1 developed by the ATR Intelligent Robotics and Communication Laboratories [58.23] (c) ROMAN, developed at the University of Kaiserslautern [58.24] (d) the Waseda DOCOMO FACE Robot No. 2 developed by Waseda University [58.25]

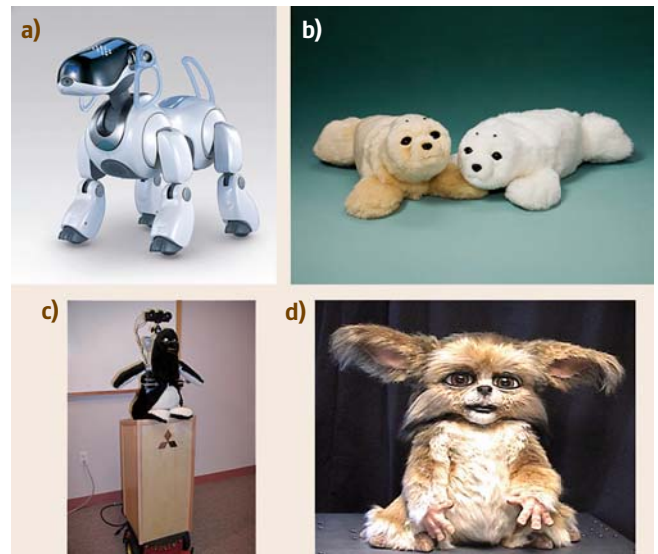


Fig. 58.3a–d Examples of social robots inspired by animals with anthropomorphic qualities (from left to right): (a) AIBO, the robotic dog previously developed by Sony [58.26], (b) Paro the therapeutic seal robot developed at AIST [58.27], (c) Mel the conversational robotic penguin developed at MERL [58.28], and (d) Leonardo developed at the MIT Media Lab [58.29]

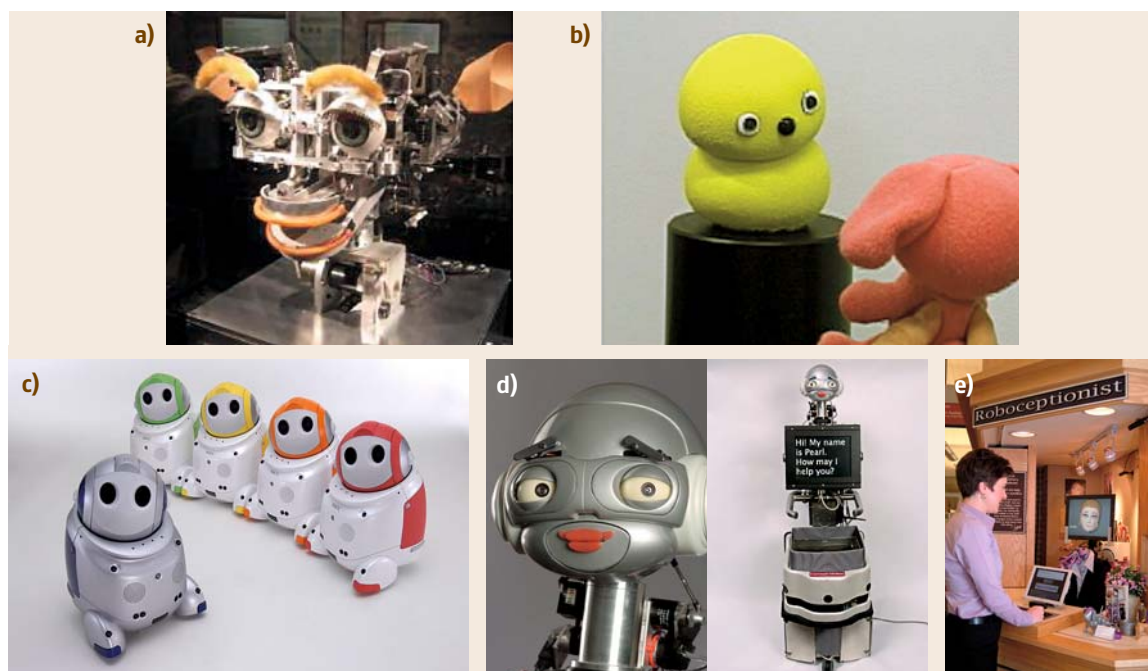


Fig. 58.4a–e Examples of social robots that are neither humanoid nor zoomorphic but capture key social attributes: (a) Kismet [58.3], (b) Keepon [58.32], (c) PaPeRo [58.35], (d) Pearl [58.34], (e) Valerie [58.33]

no overt social features like faces or eyes, but rely purely on language-based communication. Issues of proxemics on mobile social robots have also been explored such as

how a robot should approach a person [58.36], follow a person [58.37], or maintain appropriate interpersonal distance [58.38].

58.2 Multimodal Communication

Natural conversational ability is an important skill for social robots. Historically, even first-generation humanoid robots developed in the 1970s and 1980s (i.e., WABOT and WABOT-2) had primitive conversational skills modeled as simple combinations of speech input/output mappings [58.39, 40]. More recent examples are PaPeRo [58.35] and receptionist robot (ASKA) [58.41] that have conversational functions to work as information terminals.

In natural human conversation, however, people send and receive nonverbal information that supplements linguistic information. These paralinguistic cues help smooth and regulate communication between individuals.

The representative roles of paralinguistic information are as follows:

1. *Regulators*: expressions such as gestures, poses, and vocalizations that are used to regulate/control conversational turn-taking.
2. *State displays*: signs of internal state including affect, cognitive, or conversational states that improve interface transparency.
3. *Illustrators*: gestures that supplement information for the utterance. These include pointing gestures, iconic gestures,

This section focuses on conversational robots with paralinguistic communication abilities. Modern social robots employ both linguistic and paralinguistic information to perform different kinds of tasks [58.28, 29, 33].

58.2.1 Robots that Express Paralinguistic Information

In many cases, the same paralinguistic information can be conveyed through auditory or visual channels. However the characteristics of these channels have different properties. Sound has a strong property of attracting attention instantaneously. It can be used effectively for interruption, but it can also be disruptive and annoying. In addition, the timing of auditory paralinguistic cues is very strict. Overall, auditory paralinguistic signals are not suitable for continuous display. Visual paralinguistic cues are silent, the timing is a slightly less critical, and they can be used continuously. Hence, auditory and visual cues can be effectively used together to cooperatively convey the same information and emphasize it, or they can be used to convey different information simultaneously and contribute to efficiency. Thus, it is very important to choose the proper combination of modality and paralinguistic cues according to the situation. We provide examples below.

Regulatory Cues

Some of the earliest social robots displayed paralinguistic information to regulate interaction with people. Hadaly2 was the first robot to use mutual gaze as a paralinguistic cue to regulate conversation [58.42, 43]. When the robot and human achieved mutual gaze, approximated using face recognition to determine when the human's face was facing the robot, Hadaly2 expressed readiness to commence conversation by blinking its eyes. Other examples are *Kismet* [58.44] and *Leonardo* [58.29], which have implemented paralinguistic cues called *envelope displays* to regulate the exchange of speaking turns. Humans tend to make eye contact and raise their eyebrows when ready to relinquish their speaking turn, and tend to break gaze and blink when starting their speaking turn. These cues were shown to be effective in smoothing and synchronizing the exchange of speaking turns with human subjects, resulting in fewer interruptions and awkward long pauses between turns [58.3, 45].

State-Display Cues

Other robots use state-display cues whereby the face or gaze of the robot is used to indicate its conversational or cognitive state. In general, this makes the robot's internal state more transparent to a person, so they can better predict and interpret the robot's conversational state and level of understanding. ROBITA used the tightness of its facial expression to indicate readiness to engage in

conversation; a tight face was used to express conversational readiness, while a loose face communicated a lack of readiness to engage [58.5, 46].

Other state-display cues are back-channel responses. Human listeners use back-channel feedback (such as small head nods) to convey when speakers are successfully following the conversation. Response time for back-channel cues is very important because that cue is associated with the corresponding content. ROBISUKE [58.47] employed finite-state transducer (FST) technology to achieve rapid understanding of the speech signal. This allowed the robot to resolve ambiguities in meaning and prepare its own responses even when the speaker was in mid-sentence. ROBITA [58.5] provided humans with back-channel information via head nods and a tight face expression while listening.

Another back-channel signal is an expression of confusion by the listener (verbal or nonverbal). This flags the speaker to stop and try to repair the broken communication. Robots such as Leonardo and ROBITA use facial displays of confusion when speech recognition fails in order to intuitively communicate to the human that he or she should repeat their last utterance.

Similarly, gaze direction is a highly salient cue to convey the attentional focus of a robot. This is very useful if a human is trying to point out a particular object as a shared referent – such as pointing to an object before labeling it for the robot. These paralinguistic cues are useful for facilitating efficient conversational progress whereby errors and misunderstandings are identified and repaired immediately [58.45]; for instance, Leonardo's gaze behavior is used for performing envelope displays as well as for establishing shared attention with the human partner. Human subject studies have verified that Leonardo's paralinguistic cues contribute positively to the transparency of the robot's behavior and make the overall interaction more robust and efficient [58.45].

Illustrator Cues

A number of robots have implemented *illustrator* cues to direct the attention of a human [58.28, 48]. Often these robots use a variety of cues and the timing between them (such as gaze, head pose, pointing gestures, and conversational speech) to perform joint attention. In cases where a robot may be interacting with more than one person, the robot must properly take into account the location and orientation of the object, itself, and the individuals. ROBITA used such information to choose the proper gesture while considering the other people's point of view.

58.2.2 Robots that Understand Paralinguistic Information

Humans readily express paralinguistic cues when interacting with robots just as they do with people. Consequently, conversational robots must be able to recognize and properly respond to these cues as well. This is a very difficult research challenge given the wide variety, subtlety, and timing of these human cues.

Regulatory Cues

A few robots can track a human's paralinguistic cues to help regulate the conversation. The most common cues used for recognizing the end of the human's speaking turn are mutual gaze (estimated using head pose to determine when human looks back to the robot) and paused speech. As a more sophisticated example, humans frequently provide short acknowledgement utterances (e.g., 'uh-huh', 'um-hmm', 'huh', etc.) as the robot explains something. These responses are either acknowledgments or acknowledgment-like repetitions, or

ask-backs or ask-back-like repetitions. It is very difficult to distinguish these two kinds of utterances from the linguistic information as represented by the transcription of the utterance. The only way to distinguish them is by their prosody (not what, but how something is said). ROBISUKE could distinguish the utterance as either an acknowledgment or an ask-back from the prosody of utterance [58.49].

State-Display Cues

A number of robots are able to recognize and respond to state-display cues such as back-channel feedback nods, acknowledgement of an utterance, and attentional focus. One of the most robust systems for handling back channel feedback nods is Mel [58.28]. A sophisticated head nod recognition system was developed whereby the robot could successfully distinguish small feedback nods from other kinds of head nods such as those that communicate agreement. Mel used this information to determine its own nodding behavior in order to be an appropriate response for the human. In a series of human subject studies, *Sidner* et al. found these paralinguistic cues to enhance the social engagement of the robot to people [58.28]. With respect to recognizing successful or unsuccessful acknowledgement of an utterance, ROBISUKE used facial expression and prosody of the person's utterance [58.50] to make this determination. In a collaborative assembly task scenario, *Sakita* et al. [58.51] presented a robotic system that used human gaze information to deduce the human's intention of which object to operate upon next. The robot used this information to choose an appropriate cooperative action such as either taking over for a human, settling a human's hesitation, or executing simultaneously with a human.

Illustrator Cues

A number of robots are able to recognize deictic gestures of a person conveyed either through pointing gestures or head pose. For example, Leonardo is able to infer the object referent in an interaction by considering a number of factors including pointing gesture, head pose, and speech. *Brooks* and *Breazeal* [58.52] developed a deictic recognition system that enabled a robot to infer the correct object referent from correlated speech and deictic gesture. Interestingly, it was found that the accuracy of the human's pointing gesture is surprisingly poor. As a result, the deictic recognition system relies on coordinated speech and gesture information, with spatial knowledge provided by a three-dimensional (3-D) spatial database constructed by the robot using real-time vision, and a deictic spatial reasoning system. This sys-

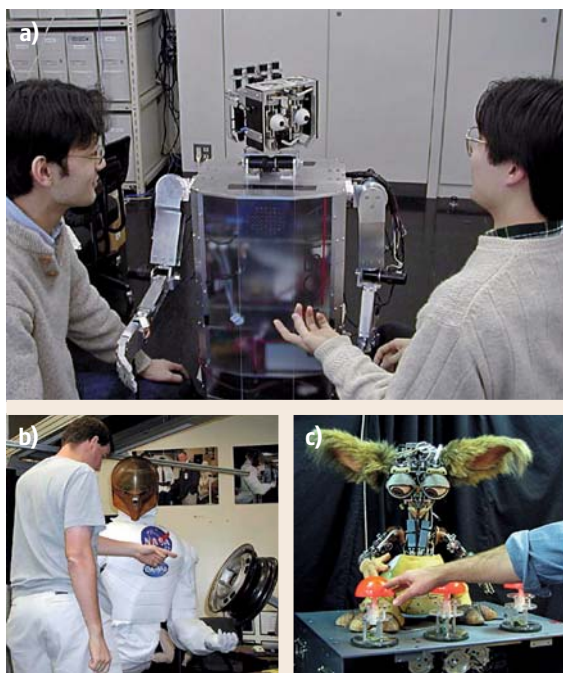


Fig. 58.5a–c Examples of conversational robots: (a) ROBITA performing group conversation; (b) Robonaut interpreting the pointing gestures of a human to determine which nut to fasten on the wheel; (c) Leonardo uses gaze and joint attention to ground the human's pointing gesture for the desired referent

tem was successfully demonstrated on the dexterous humanoid Robonaut developed at National Aeronautics and Space Administration (NASA) Johnson Space Center (JSC) (Fig. 58.5b) where the human points to and labels a set of four bolts on a wheel to be fastened in order by the robot.

58.2.3 Group Conversation

The ability to express and understand paralinguistic cues plays an important role in face-to-face conversation. The same is true for group conversation where a robot converses with two or more people. We use ROBITA, shown in Fig. 58.5a, as the illustrative example of this ability [58.46].

To continue a conversation, it is important for all conversational participants to understand who plays which role. ROBITA frames this problem with respect to information flow – to understand who is speaking to whom and when, and to determine each person’s presence as a conversational partner. During group conversation, ROBITA tries to classify the participants as the speaker or as listeners. There is only one speaker at a time and the rest are listeners. The listeners are classified into a primal listener, to whom the utterance is directed, and secondly listeners, who observe the message exchange between the speaker and the primal listener. ROBITA discerns these roles by recognizing the face direction of the speaker. The person to whom the speaker is looking is recognized as the primal listener.

To convey participation and improve its social presence in conversation, ROBITA tries to understand the conversational roles to look at the appropriate person. If the speaker faces ROBITA, then the robot recognizes itself as the primal listener and that the message is intended for itself. ROBITA faces the speaker when it is the primal listener. When ROBITA is a secondary listener, it looks at either the speaker or the primal listener.

58.2.4 Communication in Collaboration

Verbal and paralinguistic communication plays a very important role in coordinating joint action during collaborative tasks. Sharing information through communication acts is critical given that each teammate often has only partial knowledge relevant to solving the problem, each has different capabilities, and possibly diverging beliefs about the state of the task. For instance, all teammates need to establish and maintain a set of mutual beliefs regarding the current state of the task, the

respective roles and capabilities of each member, and the responsibilities of each teammate [58.53]. This is called *common ground* [58.54].

Dialog certainly plays an important role in establishing common ground. Each conversant is committed to the shared goal of establishing and maintaining a state of mutual belief with the other. To succeed, the speaker composes a description that is adequate for the purpose of being understood by the listener, and the listener shares the goal of understanding the speaker. This communication acts serve to achieve robust team behavior despite adverse conditions, including breaks in communication and other difficulties in achieving the team goals.

Humans also use nonverbal skills such as visual perspective taking and shared attention to establish common ground with others. They orient their own gaze and direct the gaze of their teammate through deictic cues such as pointing gestures in order to establish common ground. Given that visual perspective taking, shared attention, and the use of deictic cues to direct attention are core psychological processes that people use to coordinate joint action about objects and events in the world, robot teammates must be able to display and interpret these behaviors and cues when working with humans in a manner that adheres to human expectations.

Breazeal et al. [58.45] investigated the impact grounding using nonverbal social cues and behavior on task performance by a human–robot team. In a human subject experiment, participants guided Leonardo to perform a physical task using speech and gesture. The robot communicates either implicitly through behavior (such as gaze and facial expressions) or explicitly through nonverbal social cues (i.e., explicit pointing gestures). The robot’s explicit grounding acts include visually attending to the human’s actions to acknowledge their contributions, issuing a short nod to acknowledge the success and completion of the task or subtask, visually attending to the person’s attention directing cues such as to where the human looks or points, looking back to the human once the robot operates on an artifact to make sure its contribution is acknowledged, and pointing to artifacts in the workspace to direct the human’s attention toward them. Both self-reporting via questionnaire and behavioral analysis of video support the hypothesis that implicit nonverbal communication positively impacts human–robot task performance with respect to understandability of the robot, efficiency of task performance, and robustness to errors that arise from miscommunication [58.45].

58.3 Expressive Emotion-Based Interaction

Humans are fundamentally emotional beings. Consequently, human communication and social interaction often includes affective or emotive factors. To support the emotional side of human behavior, researchers are exploring affective interaction and communication between people and robots. To participate in emotion-based interactions, robots must be able to recognize and interpret affective signals from humans, possess their own internal models of emotions (often inspired by psychological theories), and be able to communicate this affective state to others.

The robot's computational model of emotion determines the robot's emotional responses according to its interactions with the external environment and its own internal cognitive-affective state. In psychology, emotional behavior depends upon many internal factors (i. e., the current emotional state, the cognitive state, current desires, physical states, etc.) These physical, cognitive, and affective states are deeply interrelated, and modulate and bias one another.

A growing number of socio-emotional robots have been designed, such as AIBO [58.26], QRIO [58.55], FEELIX [58.56], and many others. In this section, we highlight research in this area using two primary examples: Kismet and the Waseda Emotion Expression Humanoid Robot No. 4 Refined II.

58.3.1 Kismet: Inspiration from Developmental Psychology

Kismet is the first autonomous robot explicitly designed to explore socio-emotive face-to-face interactions with people [58.3]. Research with Kismet focused on exploring the origins of social interaction and communication in people, namely that which occurs between caregiver and infant, though extensive computational modeling guided by insights from psychology and ethology.

Protosocial Responses and Origins of Communication

Early infant-caregiver exchanges are heavily grounded in the regulation of emotion and its expression. Inspired by these interactions, Kismet's cognitive-affective architecture was designed to implement core protosocial responses exhibited by infants given their critical role in normal social development (see Fig. 58.6). Guided by recent psychological theories, Kismet's cognitive-affective architecture (Fig. 58.6) emphasized parallel and interacting systems of emotion and cog-

nition [58.57]. Internally, Kismet's models of emotion interacted intimately with its cognitive systems to influence behavior and goal arbitration [58.58]. Through a process of behavioral homeostasis [58.59], these emotive responses served to restore the robot's internal affective state to a mildly aroused, slightly positive state – corresponding to a state of interest and engagement in people and its surroundings that fosters learning.

One purpose of Kismet's emotive responses was to reflect the degree to which its drives and goals were being successfully met [58.60]. A second purpose was to use emotive communication signals to regulate and negotiate its interactions with people [58.61]. Specifically, Kismet utilized emotive displays to regulate the intensity of playful interactions with people, making sure to keep the complexity of the perceptual stimulus within a range the robot could handle and potentially learn from [58.62]. In effect, Kismet socially negotiated its interaction with people via its emotive responses to have humans help it achieve its goals, and satiate its drives, and maintain a suitable learning environment.

Multimodal Expressive Skills

With respect to its expressive abilities, Kismet generated a wide assortment of facial expressions with corresponding body posture to mirror its affective state. Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional space (Fig. 58.8). The basis facial postures are designed according to the componential model of facial expressions theorized by *Smith* and *Scott* [58.63] (Fig. 58.9) whereby individual facial features move to convey affective information. The three affective dimensions of the interpolation space correspond to arousal, valence, and stance. These same three attributes are used to affectively appraise the myriad of environmental and internal factors (stimuli, goals, motives, etc.) that contribute to Kismet's *emotional* state [58.58, 64].

In addition to facial expression, Kismet used an expressive vocalization system to generate a wide range of emotive utterances corresponding to joy, sorrow, disgust, fear, anger, surprise, and interest [58.65]. The robot's speech was accompanied with appropriate motor movements of the lips, jaw, and face [58.65].

Breazeal and *Aryananda* [58.66] found that even simple acoustic features (such as pitch mean and energy variance) can be used by the robot to classify the affective prosody of an utterance along valence

Antecedent conditions	Emotio	Behavior	Function
Delay, difficulty in achieving goal of adaptive behavior	Anger, frustration	Complain	Show displeasure to caregiver to modify his/her behavior
Presence of an undesired stimulus	Disgust	Withdraw, reject	Signal rejection or presented stimulus to caregiver
Presence of a threatening, overwhelming stimulus	Fear, distress	Escape	Move away from a potentially dangerous stimuli
Prolonged presence of a desired stimulus	Calm	Engage	Continued interaction with a desired stimulus
Success in achieving goal of active behavior, or praise	Joy	Display pleasure	Reallocate resources to the next relevant behavior (or to reinforce behavior in a learning context)
Prolonged absence of a desired stimulus, or scolding	Sorrow	Display sorrow	Evoke sympathy and attention from caregiver (or to discourage behavior in a learning context)
A sudden, close stimulus	Surprise	Startle response	Alert
Appearance of a desired stimulus	Interest	Orient	Attend to new, salient object
Need of an absent and desired stimulus	Boredom	Seek	Explore environment for desired stimulus

Fig. 58.6 A table summarizing the protosocial emotive responses of Kismet (adapted from *Plutchik* [58.59])

and arousal dimensions. Using these acoustic features and others suggested by *Fernald* [58.67], Kismet could recognize the affective intent in human speech as communicated through vocal prosody that corresponds to praising, soothing, scolding, and attentional bids [58.66].

58.3.2 WE-4RII: A Model of Emotion Inspired by Motion

As another example of a cognitive-affective architecture, the mental model of WE-4RII is shown in Fig. 58.10. In this section, we offer a more technical description of this model to illustrate how such mental models can be computationally implemented [58.68].

The core of this architecture is the emotion model. In dynamics, the movement of objects is described by the equation of motion. The WE-4RII mental model posits that the dynamics of mental transitions might be expressed by similar equations. Hence, the robot implements *equations of emotion* as shown in (58.1) as a second-order differential equation analogous to the equation of motion. Also, three *emotional coefficient matrixes* corresponding to the *emotional inertia*, *emotional viscosity*, and *emotional elasticity matrixes* are introduced:

$$M\ddot{E} + \Gamma\dot{E} + KE = F_{EA}, \quad (58.1)$$

where M is the emotional inertia matrix, Γ is the emotional viscosity matrix, K is the emotional elasticity matrix, and F_{EA} is the emotional appraisal.

Here, *emotional appraisal* stands for the total effect of the stimuli on mental state. By using the *equations*

of emotion, the robot can express the transient aspects of the mental state after the robot senses the stimuli from the environment. Moreover, the robot can express different reactions to the same stimuli by changing the *emotional coefficient matrixes*, and the robot can obtain complex and various mental trajectories, such as a slow reaction or an oscillated reaction [58.69, 70].

Mood Model

Human mental states are affected not only by emotion but also by mood. While emotion is characterized as a strong change that is caused in a short time, moods change more gradually over an extended time. Emotion and mood are closely related, so in this model, moods are represented by introducing the *mood vector* M_d that consists of a *pleasantness axis* and an *activation axis* shown in (58.2)

$$M_d = (M_{dP}, M_{dA}, 0). \quad (58.2)$$

Because the robot's current mental state influences the pleasantness component of the *mood vector* M_{dP} , M_{dP} is defined as the integral of the pleasantness component of the *emotion vector* as shown in (58.3). On the other hand, because the activation component of the *mood vector* M_{dA} is similar to human biological rhythm such as the internal clock [58.69, 70], the van der Pol equation (which is an equation of self-excited oscillation) is applied to describe it as (58.4).

$$M_{dP} = \int E_P dt, \quad (58.3)$$

$$\ddot{M}_{dA} + (1 - M_{dA}^2)\dot{M}_{dA} + M_{dA} = 0. \quad (58.4)$$

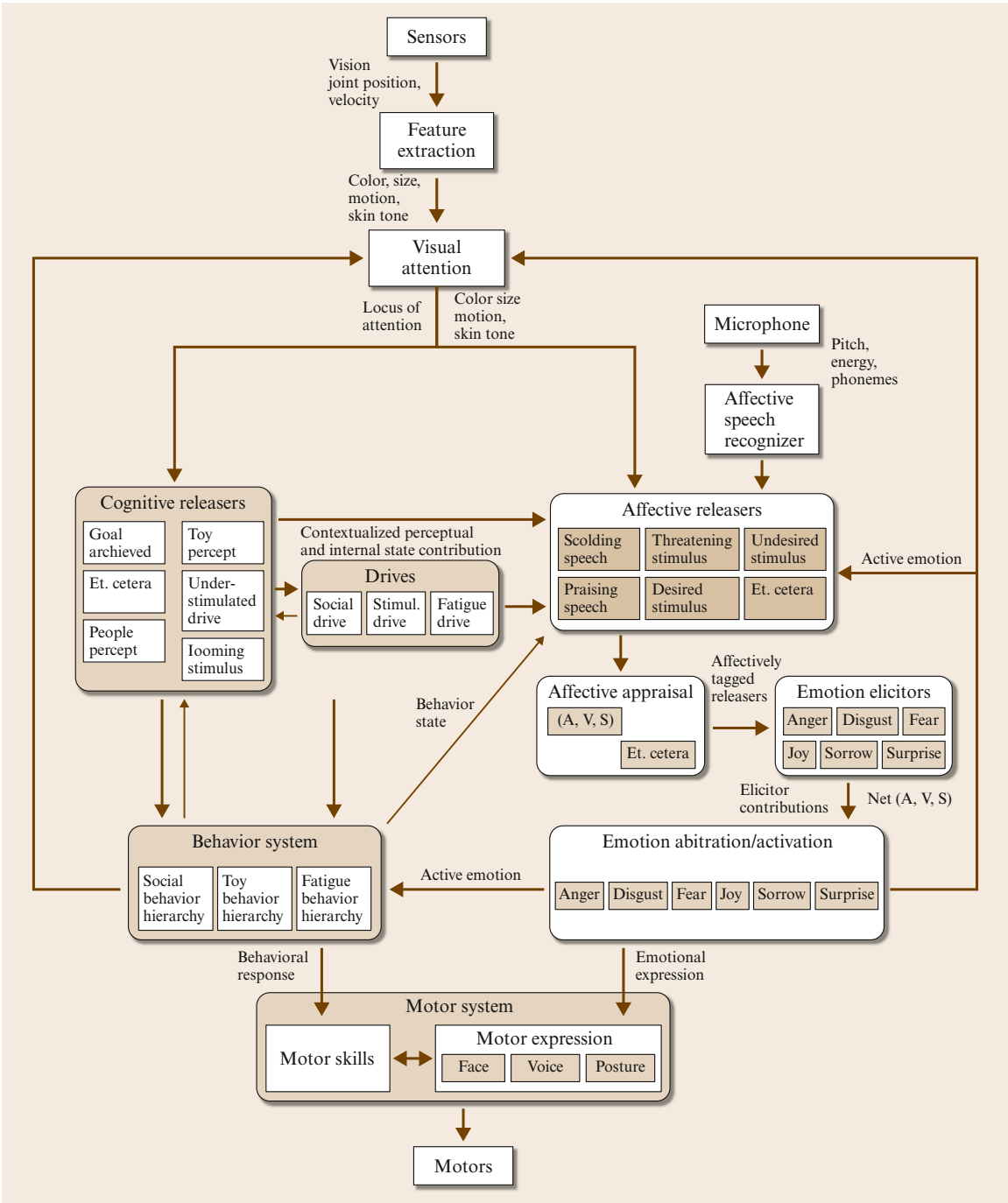


Fig. 58.7 The cognitive-affective architecture of Kismet. The gray boxes with rounded corners represent the cognitive systems responsible for perception, attention, drives, goal arbitration and execution. The white boxes represent the affective processes including affective appraisals of incoming events, basic emotive responses, and expressive motor behavior (vocalizations, facial expressions, etc.)

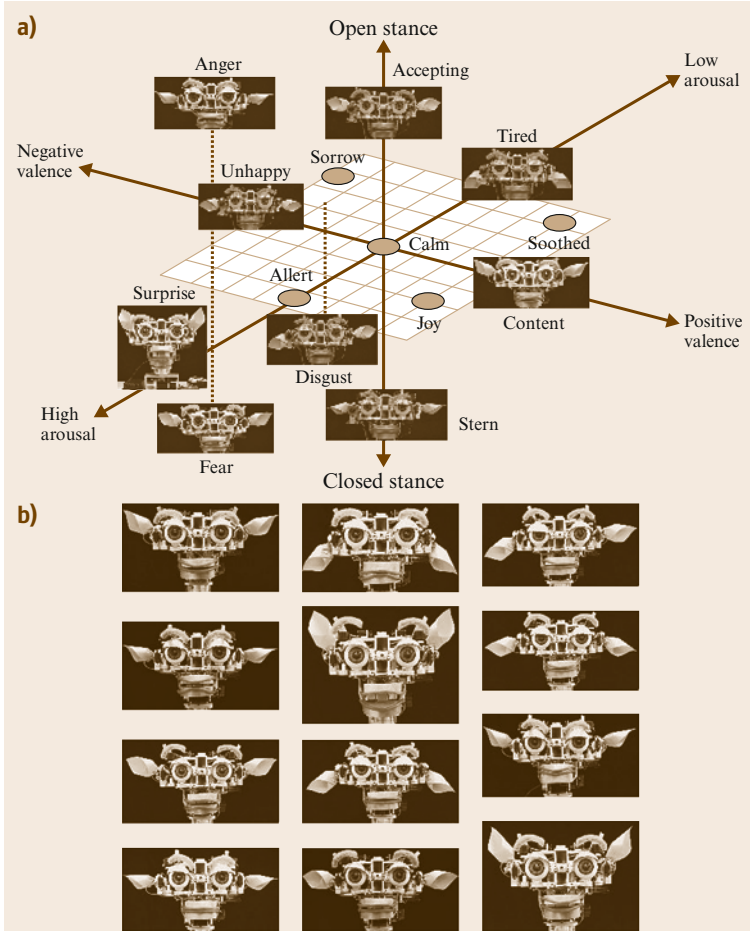


Fig. 58.8 (a) The basis facial poses of Kismet that are interpolated to generate a wide repertoire of expressions. The blend is determined according to the net affective state of the robot (arousal, valence, and stance). (b) A sampling of a range of Kismet's facial expressions from interpolation within this space of basis poses

Need model

For bilateral interaction, humans as well as robots should not only react to an individual's behavior but also behave according to internal needs. As a motivator of spontaneous behavior, a *need model* has been integrated with the mental model. Stimuli from the internal and external environment affect both the needs and the emotions of the robot. Therefore, WE-4RII has a two-layered structure of emotions and needs.

The need state of the robot is described by the *need matrix* N . The *need matrix* of WE-4RII is comprised of three kinds of needs (*appetite*, *need for security*, and *need for exploration*). However, the *need matrix* can be expanded to include other needs. The *need matrixes* N_t at time t and $N_{t+\Delta t}$ at $t + \Delta t$ are described by (58.5).

$$N_{t+\Delta t} = N_t + P_N \Delta N, \quad (58.5)$$

where P_N is the need personality matrix, and ΔN are small differences between two need states.

Equation (58.5) determines the robot need based on the stimuli from the internal and external environments. It is considered a differential equation for a continuous dynamical system. Therefore, (58.5) are named the *equations of need* [58.69, 71].

Facial Expression

WE-4RII can express seven basic emotions defined by Ekman [58.72]. To express them, a three-dimensional mental space consisting of *pleasantness*, *activation*, and *certainty* axes is defined based on psychological research ([58.73, 74]). This is represented as the *emotion vector* E described in (58.6)

$$E = (E_P, E_A, E_C). \quad (58.6)$$

Facial action								
Meaning	Eyebrow frown	Raise brows	Raise upper eyelid	(Raise lower eyelid)	Up turn lip centers	Open mouth	Tighten mouth	Raise chin
Pleasantness	Less	Disgust			More	More	Less	Less
Goal obstacle/discrepancy	More	More						
Anticipated effort	More							
Attentional activity		More	More	(More)				
Certainty		Less				More		
Novelty		More	More					
Personal agency/control		Less	Less			Less		

Fig. 58.9 A mapping of how facial features relate to underlying affective dimensions, inspired by *Smith and Scott* [58.63]. Kismet does not have lower eyelids, so we note this in parentheses

Seven emotions and their corresponding expressions are mapped into the 3-D mental space shown in Fig. 58.11.

WE-4RII determines its emotion by the *emotion vector* *E* passing through each region ([58.69, 70]).

58.4 Socio-cognitive Skills

Socially intelligent robots, however, must understand and interact with animate entities (i. e., people, animals, and other social robots) whose behavior is governed by having a mind and body. In other words, social robots need the ability recognize, understand, and predict human behavior in terms of the underlying mental states such as beliefs, intents, desires, feelings, etc. Psychology calls this ability *theory of mind* (also known as mindreading, mental perception, social common-sense, folk psychology, social understanding, among others).

This section reviews research in implementing models of human socio-cognitive skills and abilities on robots. Social robots will need a diverse repertoire of such skills to realize their full potential in daily human life – to communicate, cooperate, and learn from people in a humancentric and human-compatible manner.

For instance, social robots will need to be aware of people’s goals and intentions so that they can appropriately adjust their behavior to help us as our goals and needs change. They will need to be able to draw their attention flexibly to what we currently find of interest so that their behavior can be coordinated and information can be focused about the same thing. They need to realize that perceiving a given situation from different perspectives impacts what we know and believe to be true about it. This will enable them to bring important

information to our attention that is not easily accessible to us when we need it. Social robots will need to be deeply aware of our emotions, feelings, and attitudes to be able to prioritize what is the most important thing to do for us according to what pleases us or to what we find to be most urgent, relevant, or significant.

Furthermore, the behavior of social robots will need to adhere to people’s expectations. Namely, people will apply their theory of mind to understand the robot in terms of these mental states as well.

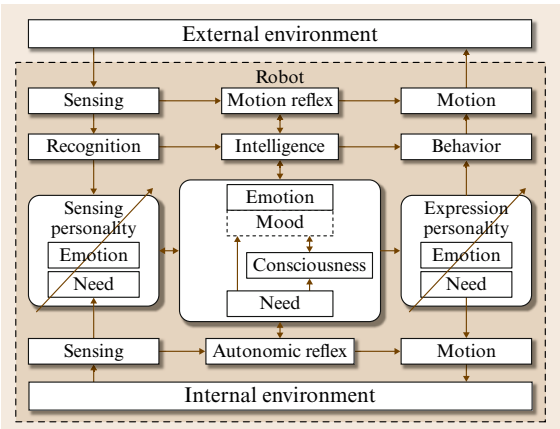


Fig. 58.10 The mental model implemented on the WE-4RII

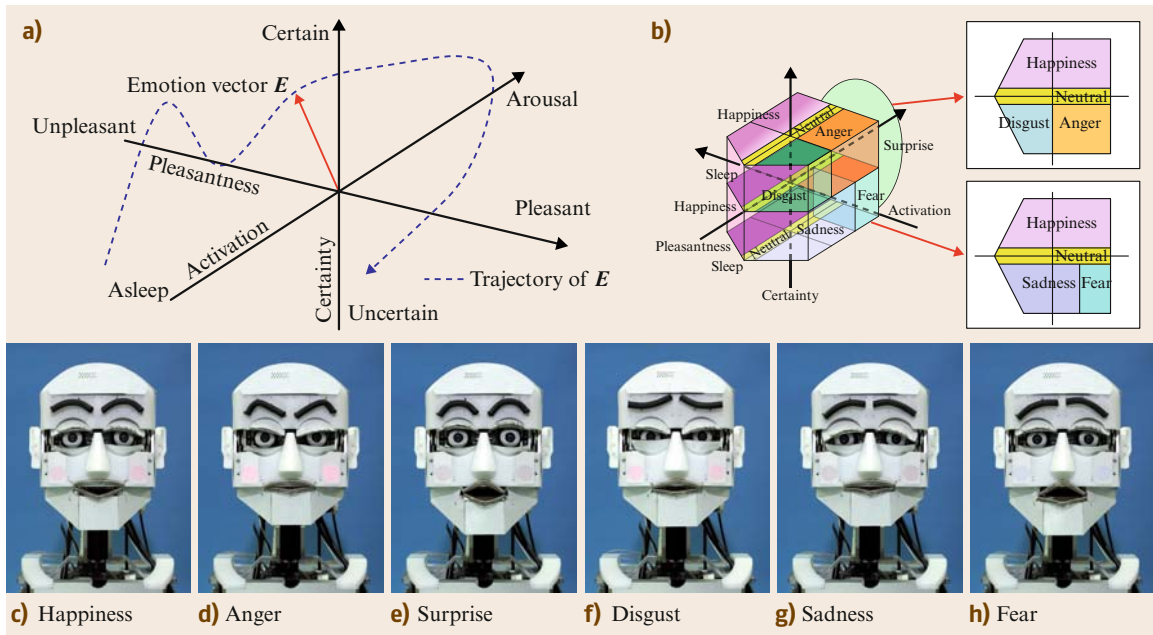


Fig. 58.11a–h The expressions of WF-4RII: (a) 3-D mental space, (b) region mapping of emotion, (c–h) seven basic facial expressions of the WE-4RII

58.4.1 Shared Attention

Scassellati's [58.75] was one of the earliest works to pose the question of how to endow robots with a theory of other minds. Inspired by the theoretical viewpoints proposed from the study of autism (believed to be a deficit of theory of mind), Scassellati implemented a hybrid model of those models proposed by Leslie [58.76] and Baron-Cohen [58.77] where shared attention is viewed to be a critical (and missing) precursor to the theory of mind competence. This hybrid model was implemented on the humanoid robot Cog. The robot was able to exhibit an assortment of social-cognitive skills such as joint attention, distinguishing an entity in the environment as either being animate or inanimate, and imitating only entities deemed to be animate.

Several researchers have explored models of joint reference, guided by insights provided by developmental psychology and autism research [58.78–80]. Normal human infants first demonstrate the ability to share attention with others at 9–12 months of age, such as following the adult's gaze or pointing gestures to the object being referred [58.77, 81]. In these works, joint attention is a learned process. For instance, the robot learns the visual motor mapping from the human's attentional cue (often using head pose as a popular indicator of what

the human is currently looking at) to the motor commands necessary to have the robot look at the same thing. This process is often bootstrapped by having the human look to where the robot initiates its gaze. In Fasel et al. [58.80], the robot learns a model of joint attention because it discovers that the human's gaze is a reliable indicator of where there is something interesting to look.

Thomaz et al. [58.82] explore attention-monitoring behavior of a robot in a social referencing interaction. In the developmental psychology literature, the ability for babies to actively monitor that others are looking at the same thing is a strong indicator of shared attention [58.83]. Social referencing is considered to be an early demonstration of shared attention because the baby looks back and forth between the novel object and the adult's emotive reaction toward that object to learn the association between the two. To implement shared attention, the robot's attentional state is modeled with two related but distinct foci: the current attentional focus (what is being looked at right now) and the referential focus (the current topic of shared focus, i.e., what communication, activities, etc. are about). Furthermore, the robot maintains a model for its own attentional state and a model for the attentional state of the human. The robot uses the heuristic of looking time upon a shared object

to infer the referent of the interaction. Once the referent has been identified, the robot monitors the attention of the human in order to associate their emotional reaction about that object to the intended target.

58.4.2 Emotional Empathy

For humans, the dynamic coupling of like minds through the actions of similar bodies is critical for acquiring human-like intuitions about the internal states of others. *Dautenhahn* [58.2] is one of the earliest works to explore *empathic* mechanisms of understanding others in social robot–robot interaction.

It is likely that emotional empathy in humans is learned, beginning in infancy. Various experiments with human adults have shown a dual affect–body connection whereby posing one’s face into a specific emotive facial expression actually *elicits* the feeling associated with that emotion [58.84, 85]. Hence, imitating the facial expressions of others could cause an infant to feel what the other is feeling, thereby allowing the infant to learn the association of observed emotive expressions of others with the infant’s own internal affective states. Other time-locked multimodal cues

may facilitate learning this mapping, such as affective speech that accompanies emotive facial expressions during social encounters between caregivers and infants.

Using a similar approach, *Breazeal* et al. [58.86] posit that a robot could learn the affective meaning of emotive expressions signaled through another person’s facial expressions, body language, and synchronized multimodal cues such as vocal prosody (see Fig. 58.12). For the robot, certain kinds of stimuli, such as pleasing or soothing tones of speech, have hardwired affective appraisals with respect to arousal and valence [58.66]. This computational model is based on the developmental findings of *Fernald* [58.67] that showed that certain prosodic contours are indicative of different affective intents in infant-directed speech. These affective intents are highly correlated with the corresponding emotive facial expression.

The tasks that couple these heterogeneous processes are face-to-face interactions and imitations. Via dual body–affect pathways, when the robot imitates the emotive facial expressions of others, it evokes the corresponding internal affective state (in terms of arousal and valence variables as described in Sect. 58.3.1) that would ordinarily give rise to the same expression during an emotive response. This is reinforced by affective information coming from the person’s speech signal.

These time-locked multimodal states occur because of the similarity in bodies and body–affect mappings, and they enable the robot to learn to associate its internal affective state with the corresponding observed expression. Thus, the robot uses its own cognitive and affective mechanisms and dual body–affect pathways as a simulator for inferring the human’s affective state as conveyed through behavior. This enables the robot to learn the association between visually observed facial expressions to the underlying affective meaning of those expressions. This is an *empathetic* approach because the robot takes on the corresponding affective state of the human in order to learn and recognize the emotional meaning of the particular facial expression.

58.4.3 Mental Perspective Taking

This section explores this empathetic, self-as-simulator approach further to address more general challenges in endowing robots with mental perspective-taking abilities. These approaches are inspired and informed by theories championed by neuroscience and embodied cognition called *simulation theory*.

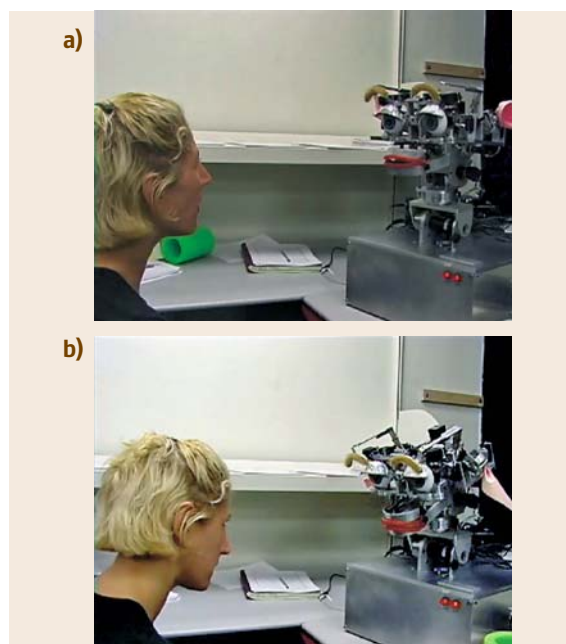


Fig. 58.12a,b Kismet and a young woman mirroring affect. Facial expression and affective tone of voice are tightly correlated: (a) mirroring interest/arousal, (b) mirroring negative affect.

Simulation Theory

Simulation theory holds that certain parts of the brain have dual use; they are used not only to generate our own behavior and mental states, but also simulate the introceptive states of the other person [58.87]. In other words, we engage in a process of perspective taking and mental simulation.

For instance, *Gallese and Goldman* [58.88] proposed that a class of neurons discovered in monkeys (called *mirror neurons*) is a possible neurological mechanism underlying both imitative abilities and simulation theory-type prediction of the behavior of others and their mental states. Further, *Meltzoff and Decety* [58.89] posit that imitation is the critical link in the story that connects the function of mirror neurons to the development of adult mindreading skills. From the field of embodied cognition, *Barsalou et al.* [58.90] present additional evidence from various social embodiment phenomena that, when observing an action, people activate some part of their own representation of that action as well as other cognitive states that relate to that action.

Mirror Systems for Recognizing Actions

Inspired by these theories and findings, *Johnson and Demiris* [58.91] employ a simulation of visual perception to recreate the visual egocentric sensory space and corresponding egocentric behavioral space of the observed agent to increase the accuracy of action recognition. This approach is based on their hierarchical attentive multiple models for execution and recognition (**HAMMER**) architecture that takes an approach inspired by mirror neurons to action recognition and imitation by directly involving the observer's motor system in the action recognition process. Specifically, during observation of another's actions, all of the observer's inverse models (akin to motor programs) are executed in parallel via simulation using forward models, and then compared to the observed action. The one that matches best is selected as being the recognized action. Perceptual perspective taking is needed to provide meaningful data for comparison. The simulated actions used by the observer during recognition must be generated as though from the point of view of the other person. They demonstrate this approach in an experiment where a robot attributes perceptions and recognizes the actions of a second robot [58.91].

Mental Perspective-Taking for Inferring Beliefs and Goals

Gray et al. [58.92] have implemented computational models of simulation-theoretic mechanisms throughout

several systems within Leonardo's cognitive architecture to enable the robot to infer beliefs and goal states of a human collaborator.

The robot reuses its belief-construction systems from the visual perspective of the human to predict the beliefs the human is likely to hold to be true given what he or she can visually observe. This enables the robot to recognize and reason about the beliefs held by a person, even when they diverge from the robot's own beliefs about the same situation.

In psychology, the ability to appreciate the divergent beliefs of another is classically demonstrated by the famous *false-belief* task. In this task, subjects are told a story with pictorial aides that typically proceeds as follows: two children, Sally and Anne, are playing together in a room. Sally places a toy in one of two containers. Sally then leaves the room, and while she is gone, tricky Anne moves the toy into the other container. Sally returns. At this point the human subject is asked "Where will Sally look for the toy?"

The robot, Leonardo, has demonstrated its ability to pass these sorts of false-belief tasks where it observes two humans playing the roles of *Sally* and *Anne* [58.93]. Within the robot's goal-directed behavior system (where schemas relate preconditions and actions with desired outcomes) motor information is used along with perceptual and other contextual clues (such as hierarchically structured task knowledge) to infer the human's goals and how he or she might be trying to achieve them (i. e., plan recognition).

58.4.4 Perspective Taking in Collaboration

By using a simulation-theoretic methodology, mental inferences made across different cognitive systems can interact in interesting and useful ways to support collaborative behavior where a robot offers its human teammate appropriate assistance.

Using Visual Perspective Taking to Resolve Ambiguous Referents

Trafton et al. [58.94] have developed and implemented visual and spatial perspective-taking abilities based on mental simulation to support human-robot interaction and collaboration. Their cognitive architecture, *polyscheme*, is designed to model how humans integrate multiple representational methods, reasoning, and planning methods to keep track of the world, including rich facilities for representing counterfactual worlds. It thus supports simulations of other people's visual perspective to reason about interac-

tions and the world from this alternate point of view.

They have demonstrated these skills in a number of experiments, such as demonstrating the robot's ability to learn how to play *hide and seek* with a person, where the robot learns what makes a good hiding place with respect to being completely occluded from the human seeker's point of view [58.95]. They have also demonstrated the usefulness of this system for a robot that solves a series of perspective-taking problems using the same frames of references and spatial reasoning abilities that astronauts do to facilitate collaborative problem solving – such as repairing a vehicle with another person that has a different vantage point [58.94]. For instance, the robot can handle egocentric requests (i. e., “hand me the cone to my right”), addressee-centric requests (i. e., “hand me the cone to your right”), or object-centered requests (i. e., “hand me the cone in front of the box”).

In Trafton et al. [58.96], a human interacts with the robot using a multimodal interface that supports speech and gesture. The robot's perspective-taking skills are used to resolve ambiguous referents that can arise when a person asks a robot perform an action in relation to an object (i. e., asking the robot to “hand me the wrench” when there are multiple wrenches to choose from). In particular, a visual occlusion in the workspace might hide another candidate wrench from the person's viewpoint but not from the robot's viewpoint (see Fig. 58.13). The robot can infer which is the intended object by taking the visual perspective of the human and applying principles of *joint salience* and *least effort*. If there still remains an ambiguity, the robot can act to resolve it by asking “which one?”

Providing Informational or Instrumental Support

Gray et al. have demonstrated the ability for the Leonardo robot to successfully infer its human partner's beliefs, desires, and intentions from real-time behavior during collaborative tasks. The shared workspace can have either visual occlusions [58.92] or can change dynamically where not all participants know of these changes [58.93]. The robot can integrate these mental state inferences to decide how best to help the person such as offering instrumental support (acting on the environment to help the human complete their goal) or provide informational support (giving relevant information that the person needs to successfully achieve his or her goal).

Consider the following scenario: a helpful robot is introduced to two people, Sally and Anne. All three

watch as Anne hides chips in a box to the left of the robot and cookies in a box to the right. Sally leaves the room, at which point Anne plays a trick on Sally by swapping the contents of the boxes and then locks both boxes with a combination lock. Anne leaves, and Sally soon returns craving the chips she saw placed in one of the boxes. Sally remembers seeing the chips placed in the left box and attempts to open it by working the combination lock. The robot has matching chips and cookies that it can give out. What should the robot do to assist Sally?

Mindreading skills play an important role in this plan recognition scenario where the robot must observe Sally in real time to infer Sally's misconception of where the chips are (Anne switched the location when Sally was out of the room), to infer what her desire is based on her behavior (Sally never explicitly said she wants the chips), and to recognize that Sally's plan for how to get the chips is actually invalid (she is trying to open



Fig. 58.13 Robonaut using visual perspective taking to disambiguate the intended referent when asked to *hand me the wrench*. The human can only see one wrench, but the robot can see both. The robot correctly hands the wrench that both can see

the wrong box). The robot has true knowledge of the situation, and must then reason about how best to help Sally get the object of her desire.

Gray et al. [58.93] combines these three kinds of mental inferences to demonstrate intention recognition with divergent beliefs for collaborative robots. Specifically, for the case of informational support, Leonardo relates its own beliefs about the state of the shared workspace to those of the human based on the visual

perspective of each. If a visual occlusion exists or an event occurs that prevents the human from knowing important information about the workspace, the robot knows to direct the human's attention to bring that information into common ground. For instance, Leonardo points to the box that actually holds the chips. For the case of instrumental support, Leonardo helps the person by directly giving the person a matching bag of chips.

58.5 Conclusion and Further Reading

In this chapter, we have presented some of the principal research trends in social robotics and human–robot interaction. We have relied heavily on examples from our own research to illustrate these trends, and have used excellent examples drawn from other research groups around the world.

From this overview, we have shown that one of the most important goals social robots as applied HRI is the creation of robots that are human-compatible and human-centered in their design. Their differences from human abilities should complement and enhance our strengths. Their similarities to human abilities, such as computationally implementing human cognitive or affective models, may help us to understand ourselves better. We expect that in the coming decades many other researchers, especially young researchers, will actively contribute to the transition from today's robots into capable robot partners of tomorrow.

For further reading, we recommend the following conference proceedings, books, articles, and journal special issues on HRI.

Annual conference proceedings for human–robot interaction:

- Proceedings of the 1st and 2nd Association for Computing Machinery (ACM)/Institute of Electrical and Electronics Engineers (IEEE) International Conference on Human–Robot Interaction (HRI 2006, HRI 2007)
- Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication Getting to Know Socially Intelligent

Robots (Ro-Man, Bellingham 2006) published by the IEEE

- Multidisciplinary Collaboration for Socially Assistive Robotics. Papers from the 2007 Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium. Technical report SS-07-07, AAAI Press, Menlo Park

Books:

- C. Breazeal: *Designing Sociable Robots* (MIT Press, Cambridge 2002)
- R. W. Picard: *Affective Computing* (MIT Press, Cambridge 1997)
- J.-M. Fellous, M. Arbib: *Who Needs Emotions: The Brain Meets the Robot*. (Oxford Univ. Press, New York 2005)

Review articles:

- T. Fong, I. Nourbakhsh, K. Dautenhahn: A survey of social robots., *Robot. Auton. Syst.* **42**, 143 – 166 (2003)

Special issues of journals on HRI:

- R. Murphy, E. Rogers (Eds.): Special Issue on Human–Robot Interaction, *IEEE Trans. Syst. Man Cybernet.* **24**(2) (2004)
- S. Kiesler and P. Hinds (Eds.): Special Issue on Human–Robot Interaction, *Human–Comput. Interact.* **9**(1-2) (2004)
- F. Laschi, C. Breazeal, C. Nakauchi (Eds.): Guest Editorial Special Issue on Human–Robot Interaction, *IEEE Trans. Robot.* **23**(5) (2007)

References

- 58.1 K. Dautenhahn: Getting to know each other – Artificial social intelligence for autonomous robots, *Robot. Auton. Syst.* **16**, 333–356 (1995)
- 58.2 K. Dautenhahn: I could be you: The phenomenological dimension of social understanding, *Cybern. Syst.* **28**, 417–453 (1997)
- 58.3 C. Breazeal: *Designing Sociable Robots* (MIT Press, Cambridge 2002)
- 58.4 H. Miwa, A. Takanishi, H. Takanobu: Experimental study on robot personality for humanoid head robot, *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* (2001) pp.1183–1188
- 58.5 T. Tojo, Y. Matsusaka, T. Ishii, T. Kobayashi: A conversational robot utilizing facial and body expressions, *Proc. IEEE SMC2000*, Vol.2 (2000) pp. 858–863
- 58.6 J. Cassell, J. Sullivan, S. Prevost, E. Churchill: *Embodied Conversational Agents* (MIT Press, Boston 2000)
- 58.7 G. Hoffman, C. Breazeal: Robots that work in collaboration with people, *Proc. AAAI Fall Symp. Intersec. Cognit. Sci. Robot.: From Interfaces Intell.* Washington (AAAI, Menlo Park 2004)
- 58.8 K. Dautenhahn, A.H. Bond, L. Canamero, B. Edmonds: *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, Multiagent Syst.Artif. Soc. Simul. Organiz., Vol.3 (Springer, Berlin, Heidelberg 2002)
- 58.9 W.R. Picard: *Affective Computing* (MIT Press, Cambridge 1997)
- 58.10 T. Fong, I. Nourbakhsh, K. Dautenhahn: A survey of social robots, *Robot. Auton. Syst.* **42**, 143–166 (2003)
- 58.11 S. Schaal: Is imitation learning the route to humanoid robots?, *Trends Cognit. Sci.* **3**(6), 233–242 (1999)
- 58.12 J. Solis, K. Chida, K. Suefujii, A. Takanishi: The development of the anthropomorphic flutist robot at Waseda University, *Int. J. Humano. Robots* **3**(2), 1–25 (2006)
- 58.13 Y. Ogura, H. Aikawa, K. Shimomura, H. Kondo, A. Morishima, H. Lim, A. Takanishi: Development of a new humanoid robot WABIAN-2, *Proc. IEEE Int. Conf. Robot. Autom.* (2006) pp. 76–81
- 58.14 H. Miwa, K. Itoh, H. Takanobu, A. Takanishi: Mechanical design and motion control of emotion expression humanoid robot WE-4R, 15th CISM-IFTOMM Symp.Robot Des.Dyn. Cont. ROM04–14 (2004)
- 58.15 F. Tanaka, J.R. Movellan, B. Fortenberry, K. Aisaka: Daily HRI evaluation at a classroom environment: Reports from dance interaction experiments, *Proc. 1st Annu. Conf. Human-Robot Interact. (HRI)* (Salt Lake City 2006) pp. 3–9
- 58.16 H. Lim, S. Hyon, S. Setiawan, S.A. Takanishi: Quasi-human biped walking, *Int. J. Inform. Educ. Res. Robot. Artif. Intell.* **24**(2), 257–268 (2006)
- 58.17 H. Lim, A. Ishii, A. Takanishi: Emotion-based biped walking, *Int. J. Inform. Educ. Res. Robot. Artif. Intell.* **22**(5), 577–586 (2004)
- 58.18 J. Solis, S. Isoda, K. Chida, A. Takanishi, K. Wakamatsu: Anthropomorphic flutist robot for teaching flute playing to beginner students, *Proc. IEEE Int. Conf. Robot. Autom.* (2004) pp.146–151
- 58.19 H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, C.M. Carrozza, P. Dario, A. Takanishi: Effective emotional expressions with emotion expression humanoid robot WE-4RII, *Proc. 2004 IEEE/RSJ Int. Conf. Intell. Robots Syst.* (2004) pp. 2203–2208
- 58.20 S. Roccella, M.C. Carrozza, G. Cappiello, P. Dario, J. Cabibihan, M. Zecca, H. Miwa, K. Itoh, M. Matsumoto, A. Takanishi: Design, fabrication and preliminary results of a novel anthropomorphic hand for humanoid robotics: RCH-1, *Proc. 2004 IEEE/RSJ Int. Conf. Intell. Robots Syst.* (Sendai, 2004) pp. 266–271
- 58.21 H. Miwa, K. Itoh, H. Takanobu, A. Takanishi: Design and control of 9-DoFs emotion expression humanoid arm, 2004 IEEE Int. Conf. Robot. Autom. (New Orleans 2004) pp.128–133
- 58.22 F. Iida, M. Tabata, F. Hara: Generating personality character in a Face Robot through interaction with human, *Proc. 7th IEEE Int. Workshop Robot Human Commun.* (1998) pp. 481–486
- 58.23 M. Shimada, T. Minato, S. Itakura, H. Ishiguro: Evaluation of Android Using Unconscious Recognition, *Proc. 6th IEEE-RAS Int. Conf. Humano. Robots* (2006) pp.157–162
- 58.24 K. Berns, J.J. Hirth: Control of facial expressions of the humanoid robot head ROMAN, *Proc. 2006 IEEE/RSJ Int. Conf. Intell. Robots Syst.* (Beijing 2006) pp. 3119–3124
- 58.25 K. Hayashi, Y. Onishi, K. Itoh, H. Miwa, A. Takanishi: Development and evaluation of face robot to express various face shape, *Proc. IEEE Int. Conf. Robot. Autom.* (2006) pp. 481–486
- 58.26 M. Fujita: On activating human communications with pet-type robot AIBO, *Proc. IEEE* **92**(11), 1804–1813 (2004)
- 58.27 K. Wada, T. Shibata, K. Sakamoto, K. Tanie: Long-term interaction between seal robots and elderly people – robot assisted activity at a health service facility for the aged, *Proc. 3rd Int. Symp. Auton. Minirobots Res. Edutain.* (2005) pp. 325–330
- 58.28 C.L. Sidner, C. Lee, C.D. Kidd, N. Lesh, C. Rich: Explorations in engagement for humans and robots, *Artif. Intell.* **166**(1–2), 140–164 (2005)

- 58.29 C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H.J. Lee Lieberman, A. Lockerd, D. Chilongo: Tutelage and collaboration for humanoid robots, *Int. J. Humano. Robots* **1**(2), 315–348 (2004)
- 58.30 M. Mori: Bukimi no tani the uncanny valley, *Energy* **7**(4), 33–35 (1970)
- 58.31 W. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, M. Wolf: Design of a therapeutic robotic companion for relational, affective touch, *Proc. 14th IEEE Workshop Robot Human Interact. Commun.* (2005) pp. 408–415
- 58.32 H. Kozima: An anthropologist in the children's world: A field study of children's everyday interaction with an interactive robot, *Proc. Int. Conf. Dev. Learn.* (ICDL) (Bloomington 2006)
- 58.33 R. Gockley, R. Simmons, J. Forlizzi: Modeling affect in socially interactive robots, *Proc. 15th IEEE Int. Symp. Robot Human Interact. Commun.* (2006) pp. 558–563
- 58.34 M.E. Pollack, S. Engberg, J.T. Matthews, S. Thrun, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, J. Dunbar-Jacob, C. McCarthy, M. Montemerlo, J. Pineau, N. Roy: Pearl: A mobile robotic assistant for the elderly, *Proc. AAAI Workshop Autom. Eldercare* (2002)
- 58.35 Y. Fujita: Personal robot PaPeRo, *J. Robot. Mechatron.* **14**(1), 60–63 (2002)
- 58.36 M.L. Walters, K. Dautenhahn, S.N. Woods, K.L. Koay: Robotic etiquette: Results from user studies involving a fetch and carry task, *Proc. 2nd ACM/IEEE Int. Conf. Human-Robot Interact.* (Washington 2007) pp. 9–11
- 58.37 R. Gockley, J. Forlizzi, R. Simmons: Natural person-following behavior for social robots, *Proc. Human-Robot Interact.* (2007) pp. 17–24
- 58.38 A.G. Brooks, C.R. Arkin: Behavioral overlays for non-verbal communication expression on a humanoid robot, *Auton. Robots* **22**(1), 55–74 (2007)
- 58.39 H. Fujisawa, K. Shirai: An algorithm for spoken sentence recognition and its application to the speech input-output system, *IEEE Trans. Syst. Man Cybern.* **SMC-4**(5), 475–479 (1974)
- 58.40 T. Kobayashi, Y. Komori, N. Hashimoto, K. Iwata, Y. Fukazawa, J. Yazawa, K. Shirai: Speech conversation system of the musician robot, *Proc. ICAR'85* (1985) pp. 483–488
- 58.41 R. Nisimura, T. Uchida, A. Lee, H. Saruwatari, K. Shikano, Y. Matsumoto: ASKA; receptionist robot with speech dialogue system, *Proc. IEEE Int. Conf. Intell. Syst. Robots* (2002) pp. 1314–1319
- 58.42 H. Kikuchi, M. Yokoyama, K. Hoashi, Y. Hidaki, T. Kobayashi, K. Shirai: Controlling gaze of humanoid in communication with human, *Proc. IROS* (1998) pp. 255–260
- 58.43 S. Hashimoto: Humanoid robots in waseda university: Hadaly2 and WABIAN, *Auton. Robots* **12**(1), 25–38 (2002)
- 58.44 C. Breazeal, A. Edsinger, P. Fitzpatrick, B. Scasselati: Active vision systems for sociable robots, *IEEE Trans. Syst. Man Cybern.* **31**(5), 443–453 (2001)
- 58.45 C. Breazeal, C. Kidd, A.L. Thomaz, G. Hoffman, M. Berlin: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork, *Proc. Int. Conf. Intell. Robot. Syst.* (2005)
- 58.46 Y. Matsusaka, T. Tojo, T. Kobayashi: Conversation robot participating in group conversation, *Trans. IEICE* **E86-D**(1), 26–36 (2003)
- 58.47 S. Fujie, K. Fukushima, T. Kobayashi: Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system, *Proc. Interspeech* (2005) pp. 889–892
- 58.48 M. Imai, T. Ono, H. Ishiguro: Physical relation and expression: Joint attention for human-robot interaction, *Proc. RO-MAN2001* (2001) pp. 512–517
- 58.49 S. Fujie, R. Miyake, T. Kobayashi: Spoken dialogue system using recognition of user's feedback for rhythmic dialogue, *Proc. Speech Prosody* (2006)
- 58.50 S. Fujie, Y. Ejiri, K. Nakajima, Y. Matsusaka, T. Kobayashi: A conversation robot using head gesture recognition as paralinguistic information, *Proc. IEEE RO-MAN 2004* (2004) pp. 158–164
- 58.51 K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, K. Ikeuchi: Flexible cooperation between human and robot by interpreting human intention from gaze information, *Proc. Int. Conf. Intell. Robot Syst.* (2004) pp. 846–851
- 58.52 A. Brooks, C. Breazeal: Working with robots and objects: revisiting deictic reference for achieving spatial common ground, *Proc. of HRI* (Salt Lake City 2006)
- 58.53 P. Cohen, H. Levesque: Teamwork, *Nous* **25**, 487–512 (1991)
- 58.54 H. Clark, D. Wilkes-Gibbs: Referring as a collaborative process, *Cognition* **22**, 1–39 (1986)
- 58.55 F. Tanaka, K. Noda, T. Sawada, M. Fujita: Associated emotion and its expression in an entertainment robot QRIO, *Proc. 3rd Int. Conf. Entertain. Comput.* (Eindhoven 2004) pp. 499–504
- 58.56 L. Canamero: <http://www.felix-growing.org/>, February 5 (2007)
- 58.57 C. Breazeal: Function meets style: Insights from emotion theory applied to HRI, *IEEE Trans. Syst. Man Cybern. Part C* **34**(2), 187–194 (2003)
- 58.58 C. Breazeal: Emotion and sociable humanoid robots, *Int. J. Human Comput. Interact.* **58**, 119–155 (2003)
- 58.59 R. Plutchik: Emotions: A general psychoevolutionary theory, In: *Approaches to Emotion*, ed. by K. Sherer, P. Elkman (Erlbaum, Hillsdale 1984) pp. 197–219
- 58.60 C. Breazeal, B. Scasselati: Infant-like social interactions between a robot and a human caregiver, *Adaptive Behavior* **8**(1), 47–72 (2000)

- 58.61 C. Breazeal: Early experiments using motivations to regulate human-robot interaction, Proc. 1998 Amer. Assoc. Artif. Intell. (AAAI98) Fall Symposium on Emotional and Intelligent: the Tangled knot of Cognition, Techn. Rep., Vol. FS-98-03 (Orlando 1998) pp. 31-36
- 58.62 C. Breazeal, A. Edsinger, P. Fitzpatrick, B. Scasselati: Active vision systems for sociable robots, IEEE Trans. Syst. Man Cybern. **31**(5), 443-453 (2001)
- 58.63 C. Smith, H. Scott: A componential approach to the meaning of facial expressions. In: *The Psychology of Facial Expression*, ed. by J.J. Russell Fernandez-Dols (Cambridge Univ Press, Cambridge 1997) pp. 229-254
- 58.64 C. Breazeal: Regulation and entrainment for human-robot interaction, Int. J. Exp. Robot. **21**(10-11), 883-902 (2003)
- 58.65 C. Breazeal: Emotive qualities in lip synchronized robot speech, Adv. Robot. **17**(2), 97-113 (2003)
- 58.66 C. Breazeal, L. Aryananda: Recognizing affective intent in robot directed speech, Auton. Robots **12**(1), 85-104 (2002)
- 58.67 A. Fernald: Intonation and communicative intent in mother's speech to infants: Is the melody the message?, Child Develop. **60**, 1497-1510 (1989)
- 58.68 K. Itoh, H. Miwa, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M.C. Carrozza, P. Dario, A. Takanishi: Behavior model of humanoid robots based on operant conditioning, Proc. 2005 IEEE-RAS Int. Conf. Humano. Robots (Tsukuba 2005) pp. 220-225
- 58.69 H. Miwa, K. Itoh, H. Takanobu, A. Takanishi: Development of mental model for humanoid robots, 15th CISM-IFTOMM Symp. Robot Des. Dyn. Contr., Vol. ROM04-10 (Montreal 2004)
- 58.70 H. Miwa, T. Okuchi, K. Itoh, H. Takanobu, A. Takanishi: A new mental model for humanoid robots for human friendly communication: Introduction of learning system, mood vector and second order equations of emotion, Proc. IEEE Int. Conf. Robot. Autom. (2003) pp. 3588-3593
- 58.71 H. Miwa, K. Itoh, K. Ito, H. Takanobu, A. Takanishi: Introduction of the need model for humanoid robots to generate active behavior, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (2003) pp. 1400-1406
- 58.72 P. Ekman, W.V. Friesen: The repertoire of nonverbal behavior, Semiotica **1**, 49-98 (1969)
- 58.73 J.A. Russell, M. Bullock: Multidimensional scaling of emotional facial expressions: Similarity from preschool to adults, J. Pers. Social Psychol. **48**, 1290-1298 (1985)
- 58.74 C.A. Smith, P.C. Ellsworth: Patterns of cognitive appraisal in emotion, J. Pers. Social Psychol. **48**, 813-838 (1985)
- 58.75 B. Scasselati: Foundations for a Theory of Mind for a Humanoid Robot. Ph.D. Thesis (MIT, Department of Computer Science and Electrical Engineering 2000)
- 58.76 A. Leslie: How to acquire a representational theory of mind. In: *Metarepresentation: A Multidisciplinary perspective*, ed. by D. Sperber (Oxford Univ Press, Oxford 1994) pp. 197-223
- 58.77 S. Baron-Cohen: Precursors to a theory of mind: Understanding attention in others. In: *Natural Theories of Mind*, ed. by A. Whiten (Blackwell, Oxford 1991) pp. 233-250
- 58.78 B. Scasselati: Mechanisms of shared attention for a humanoid robot, Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium (AAAI Menlo Park, 1996)
- 58.79 H. Kozima: Attention-sharing and behavior-sharing in human-robot communication, IEEE Int. Workshop Robot Human Commun. (1998) pp. 9-14
- 58.80 I. Fasel, G.O. Deak, J. Triesch, J. Movellan: Combining embodied models and empirical research for understanding the development of shared attention, Intl. Conf. on Development and Learning (2002) pp. 21-27
- 58.81 G. Butterworth: The ontogeny and phylogeny of joint visual attention. In: *Natural Theories of Mind*, ed. by A. Whiten (Blackwell, Oxford 1991) pp. 223-232
- 58.82 L.A. Thomaz, M. Berlin, C. Breazeal: An embodied computational model of social referencing, Proc. of 14th IEEE Workshop Robot Human Interact. Commun. (Nashville 2005)
- 58.83 S. Feinman: Social referencing in infancy, Merrill-Palmer Quarterly **28**, 445-470 (1982)
- 58.84 F. Strack, L. Martin, S. Stepper: Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis, J. Personal. Social Psychol. **54**, 768-777 (1988)
- 58.85 M.P. Niedenthal, W.L. Barsalou, P. Winkielman, S. Krauth-Gruber, F. Ric: Embodiment in attitudes, social perception, and emotion, Personal. Social Psychol. Review **9**(3), 184-211 (2005)
- 58.86 C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, B. Blumberg: Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots, Artif. Life **11**(1-2), 1-32 (2005), ed by. L. Rocha and F. Almedia e Costa
- 58.87 M. Davies, T. Stone: *Mental Simulation* (Blackwell, Oxford 1995)
- 58.88 V. Gallese, A. Goldman: Mirror neurons and the simulation theory of mind-reading, Trends Cognit. Sci. **2**(12), 493-501 (1998)
- 58.89 A. Meltzoff, J. Decety: What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience, Philos. Trans. R. Soc. Lond. B Biol. Sci. **358**, 491-500 (2003)
- 58.90 L.W. Barsalou, P.M. Niedenthal, A. Barbey, J. Rupert: Social embodiment. In: *The Psychology of Learning and Motivation*, Vol. 43, ed. by B. Ross (Academic Press, San Diego 2003) pp. 43-92

- 58.91 M. Johnson, Y. Demiris: Perceptual perspective taking and action recognition, *Int. J. Adv. Robot. Syst.* **2**(4), 301–308 (2005)
- 58.92 J. Gray, C. Breazeal, M. Berlin, A. Brooks, J. Lieberman: Action parsing and goal inference using self as simulator, *Proc. 14th IEEE Workshop Robot Human Interact. Commun.* (Nashville 2005)
- 58.93 J. Gray, M. Berlin, C. Breazeal: Intention recognition with divergent beliefs for collaborative robots, *Proc. AISB 2007: Symp. Mindful Environ.* (Newcastle upon Tyne 2007)
- 58.94 J.G. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, A. Schultz: Enabling effective human-robot interaction using perspective-taking in robots, *IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans* **35**(4), 460–470 (2005)
- 58.95 G. Trafton, C.A. Shultz, D. Perzanowski, W. Adams, M. Bugajska, N. Cassimatis, D. Brock: Children and robots learning to play hide and seek, *Proc. 1st Annu. Conf. Human-Robot Interact.* (Salt Lake City 2006)
- 58.96 G. Trafton, A. Schultz, M. Bugajska., F. Mintz: Perspective-taking with robots: Experiments and models, *Proc. of 14th IEEE Workshop Robot Human Interact. Commun.* (Nashville 2005)