In: Proceedings Third Cognitive Technology Conference CT'99, August, San Francisco.

# ROBOTS AS SOCIAL ACTORS: AURORA AND THE CASE OF AUTISM

**Kerstin Dautenhahn**

*Department of Cybernetics, University of Reading, Whiteknights PO BOX 225, Reading RG6 6AY, UK*
*MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139, USA*

*Email: K.Dautenhahn@reading.ac.uk*
*URL: http://www.cyber.rdg.ac.uk/people/kd/WWW/home.html*

## INTRODUCTION: THE LIFE-LIKE AGENT HYPOTHESIS

This paper discusses the role of predictability and control in robot-human interaction. This involves the central question whether humans are good models for synthetic (social) agents. Design issues based on cognitive accounts towards robot-human interaction are discussed with respect to the author's recent work on building interactive robotic systems as remedial tools (teaching devices) for children with autism, a project which crucially requires a careful analysis of human-robot relationships, and which cannot rely on the 'natural' tendency of humans to be interested in agents that appear life-like and social, a factor which is currently exploited extensively in the believable agents research community. The paper will argue that using humans as models for creating believable technology (e.g. making the robot life-like) is not a universal solution for designing robotic social actors.

Often, research into social agents seems to be based on what can be called the 'life-like agent hypothesis' and can be phrased as follows:

*Artificial social agents (robotic or software) which are supposed to interact with humans are most successfully designed by imitating life, i.e. making the agents mimic as closely as possible animals, in particular humans. This comprises both 'shallow' approaches focussing on the presentation and believability of the agents, as well as 'deep' architectures which attempt to model faithfully animal cognition and intelligence. Such life-like agents are desirable since*

1. *The agents are supposed to act on behalf of or in collaboration with humans; they adopt roles and fulfill tasks normally done by humans, thus they require human forms of (social) intelligence.*
2. *Users prefer to interact ideally with other humans and less ideally with human-like agents. Thus, life-like agents can naturally be integrated in human work and entertainment environment, e.g. as assistants or pets.*
3. *Life-like agents can serve as models for the scientific investigation of animal behaviour and animal minds.*

Certainly argument (3) is valid and we will not discuss it further. This article points out that arguments (1) and (2) which may appear straigthforward and convincing at first, actually gloss over difficult issues which need careful analysis. To start with, we discuss consequences of the life-like agent hypothesis, in particular correspondences between animals and artifacts, focussing on agents.

Animals are embodied, living beings, which creates strong constraints on what they can do and how humans can use them. Even genetic engineering cannot produce arbitrary bodies which can survive. Animal bodies are adapted to a particular ecological niche consisting of all abiotic and biotic factors influencing an animal's life. Thus, taking the metaphor of life-like behaviour seriously means to abandon the ideal of universal building blocks which can be assembled in arbitrary order in order to build machines according to our specification. Shaping of animal behaviour and cognition is constrained by the limits of behavioural and cognitive capabilities. For example, despite long efforts dogs or chimpanzees cannot 'speak' and humans cannot fly without the help of machines.

Animals are autonomous beings and cannot be totally controlled by humans in the same way as we control a car or a mobile phone. Animals can be educated, trained; their behaviour can be shaped, but not in the same way as we can shape a stone to result in a beautiful sculpture. Creating complex life-like agents means to accept the risk of partially losing control of the agent's behaviour.

Animal behaviour is often difficult to predict in particular in highly individualised societies like human societies. According to the *Social* or *Machiavellian Intelligence Hypothesis* [Byrne & Whiten, 1988; Byrne 1995] humans have large brains and high intelligence because of their need to manage and predict complex social dynamics. Humans show a 'natural' tendency to anthropomorphise (attributing human characteristics, in particular mental concepts, to non-human entities) and to adopt an 'intentional stance' (treating a system as if it has intentions, irrespective of whether it actually does). This has been discussed to be adaptive in the sense that animating clouds or bisons might help to predict their behaviour [Dennett, 1987; Watt, 1997]. Our tendency to anthropomorphise might be due to the fact that our prediction abilities were strongly linked to (if not originating from) the social domain and are even today probably most effective there. Anthropomorphising, although it might originate in the social domain, can be easily applied to various objects, as children's pretend play demonstrates. Making agents life-like can support this natural process of anthropomorphising, but additionally it 1) evokes expectations about the behavioural and cognitive complexity of the agent compared to its natural model (disappointment of these expectations can severely reduce the believability of the agent, e.g. see [Dryer, 1999]), and 2) limits the scope of possible interpretations of the agent by the user, depending on the user's individual personality, preferences and social/cultural context. Due to the latter effect the more life-like an agent becomes, the more specific becomes the 'meaning' it conveys and the range of possible interpretations and usages: a child can use a wooden stick to stand for a knight's sword, an Indian's arrow, or it can use it as a tool for reaching a kite which got entangled in a tree. Thus, the practical and imaginative 'function' of a wooden stick is much less restricted than e.g. the exact (plastic) replication of Excalibur. The meaning of an object is created in the mind; it is not a property of the object itself or the visual image on the retina. Thus, non-specificity with respect to appearance or behaviour of an agent could be less life-like, but could enhance the usefulness of an agent.

In the following we will discuss in more detail the issues of control and predictability in human-agent interaction, focussing on the human-robot case. A particular application area (rehabilitation of children with autism) will be described to exemplify the issues and outline some difficulties in a research direction which cannot rely on the human tendency to anthropomorphise.

## PREDICTABILITY OF HUMANS AND ROBOTS

Issues which play a role in human-human interaction can be useful guidelines in designing socially intelligent agents which are supposed to interact with human users ([Dautenhahn 1997b], [Dautenhahn 1998b]).

Many approaches are studying how robots or software agents learn from users and mimic their behavior, using machine learning techniques, like programming by demonstration [Cypher 1993]. Such techniques are resulting in products which reflect the individual behaviour or preferences of users. The degree of mimicking can range from extracting single individual features to more complex forms of imitation. Imitation is not only an efficient social learning mechanism (see overview in [Heyes & Galef, 1996]), but is used as a mechanism by which infants get to know other persons as objects which imitate and can be imitated, that is, objects which pass the 'like me' test [Meltzoff, 1995]. Recent research on imitating robots shows the relevance of imitation for social learning and communication between agents, see [Hayes & Demiris 1994], [Gaussier et al 1998], [Billard & Dautenhahn 1998]. Thus, imitation and social learning mechanisms can make socially intelligent agents more 'like us' and make them individuals. Other recent work on software agents discusses how to incorporate emotions and personality in agents, e.g. [Loyall & Bates 1997], [Trappl & Petta 1997].

Thus, there is a strong tendency toward creating agents, which possess personality, expressiveness, emotional responses and other traits, which are characteristics of humans. Users tend to treat computers like humans, using similar criteria to decide on whether they like them or not [Dryer 1999]. Humans are also known to naturally anthropomorphise artificial and natural agents inhabiting their world [Foner 1997], as an efficient adaptation, i.e. in order to predict actions, identify dangers, and to have fun. But are humans necessarily good models for synthetic agents?

Shneiderman argues that humans have a strong desire to be in control and usually seek predictable systems [Shneiderman 1992, p. 86]. Humans can be good models for synthetic agents in application areas where human-style interaction benefits the acceptance of a system, e.g. in domains which are primarily social, like creating artificial actors in virtual environments like MUDs, where the main purpose of the domain is entertainment. However, in other domains human qualities in synthetic agents might make a system unnecessarily computationally complex, and might hinder an efficient use of the system.

Predictability of others' behaviours plays an important part in human society. The need to cope with complex social relationships, to acquire and manage social knowledge in order to predict the behavior of group members is, according to the S*ocial Intelligence Hypothesis*, a decisive factor in the evolution of human intelligence (cf. previous section). Primate societies require constant monitoring, re-assessing and re-learning of relationships and group structures. What happens if you select fifty potentially dangerous animals, like lions or chimps that do not know each other, and let them travel together in a confined space? It is likely that severe conflict will arise, resulting in injuries and deaths. However, this is exactly what millions of commuters using public transportation do twice a day. Human (social) intelligence allows to cope even with extreme and highly unnatural social situations, as humans are experts in reading cues which give hints on intentions and goals of others, and they use these skills for predicting the behaviour of fellow humans. On the one hand knowledge about conventions, norms and other control mechanisms can help to interpret correctly others' behaviours, in particular in large groups of humans. On the other hand empathic understanding [Dautenhahn 1997] can help to handle individual, difficult cases. However, empathy requires identifying another agent as similar, and the degree of similarity influences our understanding. Experiments by Eddy et al. [Eddy et al., 1993] demonstrate that the higher people judge the similarity of animals to themselves, the higher is the degree to which people engage in anthropomorphism.While invertebrates, fish, amphibians and reptiles are not very likely to be subject to anthropomorphic projection, dogs, cats and primates are, in particular human primates.

Are humans able to predict the behaviour of artifacts, like software agents or robots? The interface can be designed to give clear indication about functionality and built-in goals of the agent, so that the user knows what s/he can expect. Believable-character technology can give the illusion that the agent is a sentient and intentional being. Thus, from the viewpoint of the human user, humanoid robots might be perceived as more 'similar' to humans than non-humanoid robots. Moreover, imaginative skills of humans can compensate for technological deficits. Children's pretend play with dolls and stuffed animals shows that to be perceived as 'life-like' does not necessarily require mimicking nature in all its complexity. However, children are not mistaken about the nature of a doll; they know that it is a doll, although they might have an emotional attitude towards it which resembles relationships to humans. We can expect that life-like agent might give the illusion of life but are not mistaken for life. Thus, building life-like agents is not a necessary condition for building agents, it depends on the context and purpose. While humans are good at behaving unpredictably, agents (hardware and software) are on the other hand very good at producing predictable behaviour. The strength of software and robotic agents to produce rigid, repetitive actions was their primary usage, e.g. computer programs for 'number crunching' (computations), or robots in factories assembling products according to a rigid scheme. When is it desirable to create agents that act like humans, e.g. which show complex and difficult to predict behaviour?

**Predictable**

| Humans | | Agents | |
|---|---|---|---|
| Pos. | Neg. | Pos. | Neg. |
| reliable | boring<br>unremarkable | non-distracting | boring |

**Unpredictable**

| Humans | | Agents | |
|---|---|---|---|
| Pos. | Neg. | Pos. | Neg. |
| creative | unreliable | interesting<br>attention-catching | frightening<br>intrusive |

**Figure 1.** Comparison of predictable and unpredictable agents.

Figure 1 compares attitudes towards predictable /unpredictable humans/agent which are likely to occur. Humans are quite tolerant towards unpredictable human behaviour like sudden changes in movement. If we are walking along a street during daylight and another person is walking behind us and suddenly starts to run then we usually do not feel bothered. Although the behaviour itself was not predicted, we easily have different potential explanations at hand, since we are familiar with human behaviour and the agent who was walking behind us was a human. Despite local and cultural differences, by being member of the same species we have good chances to explain human behaviour. Humans show a great variety of different behaviours and ways how they interact with the world. Their behaviour can usually be predicted by other humans quite well, based on knowledge of social rules and conventions, and their skills of anthropomorphism and 'mindreading'. A powerful skill is the projection of what one knows about human society and individual experience with the environment from the first-person perspective (the intra-agent perspective), to another person (the inter-agent perspective, mapping of the 'self' to the 'other', [Nehaniv, 1999]). Humans are much less predictable by people with autism which lack sophisticated skills of mindreading and empathy (see next section).

Imagine the same situation we described above, only this time with a robot walking behind us along the street and the robot suddenly starts to run and quickly moves up closer from the back. The robot is a humanoid, being able to express internal states, showing appropriate 'being in a hurry' facial and bodily expressions which look very human-like. However, this 'expressiveness' was specifically designed by humans for robots which exist in human-inhabited spaces, it is not based on a human-like way of experiencing the world. Thus, it can be assumed that the threshold which makes things appear frightening to us is much lower for robots than for humans. It is a natural threshold of 'suspicion', of not being biological creatures, which separates agents and humans. Life-like agents might appear like humans but they *are* not human. Robots might be our 'mind children' (see arguments in [Moravec, 1988]) but they are not phylogenetically related to us, they are not part of the 'tree' (or better: bush) of life, they do not share with us basic principles of physiology, development and inheritance, and even structural similarity of humanoid robots only holds on a superficial level (humanoid robots do not possess bones, muscles and joints based on the same principles as human anatomy). In future humans might get increasingly familiar with humanoid intelligent machines, but at present they are not part of our daily life; they are mostly part of the world of science fiction, and research laboratories. In the same way as multi-cultural co-existence in human societies is a long and difficult process, we can expect that the process of cultural integration of humanoid robots into human society will be a gradual and difficult process.

In certain application domains like entertainment, life-like believable agents can be very successful. However, believable agent technology is not a universal solution to all problems of human-machine interface design. In other applications, i.e. where stereotypical and routine work is dominating, interaction with life-like agents might be less efficient and also more annoying than conventional technology, since understanding their behaviour requires a lot of 'mental' effort. Making robots more and more interesting and life-like does not necessarily increase their usability for humans. While a socially pro-active human might be perceived as persistent (positive) or pushy (negative), a robot with the same traits is likely to appear intrusive and scary.

A recent example of life-like, believable robot technology is the platform KISMET ([Breazeal 1998]) which possesses eyebrows, ears, a mouth etc. and is able to express 'emotions' depending on the way a human interacts with the robot. This project demonstrates that humans possess specific social skills that can be exploited in designing socially intelligent agents. An underlying assumption is that the human user is motivated and interested to interact with the robot, only then interesting interactions can emerge. KISMET's life-like behaviour and appearance prompts the user to get engaged in infant-caretaker interactions, where the human interest drives the interactions and provides a 'benign' environment for the robot. The next section discusses an application where we cannot rely on this assumption.

## A ROBOT AS A SOCIAL MEDIATOR: AUTISM AND THE AURORA PROJECT

### BACKGROUND

In 1998 we started the project AURORA (AUtonomous RObotic platform as a Remedial tool for children with Autism) which investigates how an autonomous mobile robot can be used as a remedial tool in order to encourage children with autism to take initiative and use the robotic 'toy' to become engaged in a variety of different actions. Figure 2 shows the robot which we are currently using in the AURORA project.

People with autism are a very heterogeneous group and it is difficult to list defining symptoms. Depending on what is counted as 'autism', rates of occurrence are given which range between 5-15 in 10000. Instead of a physical handicap which prevents people from physically interacting with the environment, people with autism have great difficulty in making sense of the world, in particular the social world. Autism occurs in different degrees of severity, it can, but need not be accompanied with learning disabilities. The most common characteristics of autism are:

1) inappropriate social interactions, qualitatively impaired social relationships, inability to relate to others in a meaningful way, impaired capacity to understand other's feelings or mental states
2) impairment of social communication skills (e.g. understanding gestures, facial expressions, tone of voice) and fantasy, limited range of imaginative activities
3) significantly reduced repertoire of activities and interests (stereotypical behavior, fixation to stable environments).

At the higher functioning end of the autism spectrum we find people with Asperger Syndrome. Some of them manage to live independently as adults and to succeed in their profession, but only by learning and applying explicit rules in order to overcome the 'social barrier' [Grandin, 1995; Grandin & Scariano, 1996]: instead of picking up and interpreting social cues 'naturally' they can learn and memorise rules about what kind of behaviour is socially appropriate during interaction with non-autistic people. Autism is not, as has long been assumed in public, a voluntary decision to retract from the world: people with autism do not have the choice to live socially or not, the decision has been made for them. Only personal determination and intelligence gives some of them the chance to appear successfully 'socially adapted'.

Two different viewpoints exist on how to connect the autistic with the non-autistic world: either efforts are undertaken to teach people with autism the skills they need to survive in the world of 'normal' people[*], or it is suggested that they might be happier living separately in a world specifically designed for them. From all what we know about the way individuals with autism

---

[*] The author generally rejects the terms 'normality' and 'normal people', but she uses these terms in this article for the purpose of communicating the perceived difference between people who have been diagnosed autistic and those who have not.

feel (see books written by Temple Grandin and other people with autism), they are very aware of their ‘being different’ from other people, and express the wish to be part of the ‘world outside’. Accepting the differences, empowering people with autism, and linking their world with the world non-autistic people are living in poses many challenges. In order to understand people with autism we have to understand better the causes of autism, and have to find ways to empower them, e.g. by means of computer and robotic technology, so that they have the choice of whether and to what extend they want to connect to the world of non-autistic people.

Using a robot as a remedial toy takes up the challenge of bridging the gap between the variety and unpredictability of human social behaviour (which often appears frightening to children with autism) and the predictability of repetitive and monotonous behaviour which children with autism prefer and which can be performed by mobile robots.

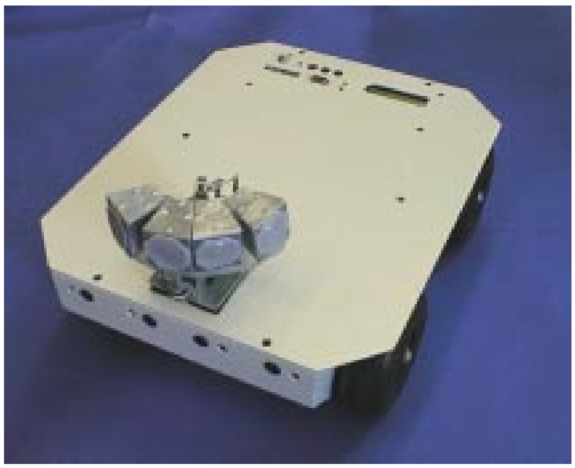

**Figure 2.** Robot used in the AURORA project, donated by *Applied AI Systems, Inc.*

## EXPLANATIONS AND METHODS OF TREATMENT

A variety of explanations of autism have been proposed. During the 50s and early 60s a Freudian theory of autism was popular which stated that children become autistic due to a lack of affection and emotional ‘warmth’ by their parents. Fortunately (in terms of negative consequences this had for the lives of many parents and their children) this explanation has been abandoned. Currently widely discussed is the ‘theory of mind’ model which is conceiving autism as a cognitive disorder ([Baron-Cohen et al 1985]). Other explanations focus on the interaction dynamics between child and caretaker ([Hendriks-Jansen 1997]). Similarly, a lack of empathic processes is suggested which prevent the child from developing ‘normal’ kinds of social action and interaction [Dautenhahn 1997]. Other evidence suggests that not impairments of mental concepts, but rather disorders of executive functions, namely functions which are responsible for the control of thought and action, are primary to autistic disorder [Russell 1997].

A variety of approaches are addressing the communication problems of children with autism, among them psychoanalytic, behavioural and developmental/psycholinguistic approaches, see overview in [Watson et al., 1989]. The psychoanalytic approach by Bruno Bettelheim [Bettelheim, 1967] assumed that children with autism use language to express indirectly and symbolically internal conflicts regarding self-differentiation. Most professionals have turned to other methods of intervention. Language disorders of children with autism are now widely understood as being based on cognitive, social and linguistic deficits.

Developmental/psycholinguistic approaches studied in detail the particular language skills and their order in development, in comparison with the development of linguistic skills in normal children, and with social and cognitive skills. Semantic and pragmatic aspects of language were addressed in instructional programs. Behavioural approaches, e.g. in the work of O. I. Lovaas [e.g. Lovaas, 1977] and his associates used operant conditioning techniques in structured therapy sessions which "…would begin with teaching a child to sit in a chair for a specified length of time, to make eye contact on command, and to imitate motoric behaviors of the trainer. Then the child would be taught to imitate vocal sounds, verbal sounds, and words, and to demonstrate receptive understanding of words by choosing an appropriate object or picture from among a set in response to verbal cues from a trainer (for example, "Dog. Touch dog.") Then the child would be taught to label objects or pictures or some feature of them, in response to a verbal cue (for example, "What is this?" or "Where is the ball?"). A child who advance further would be taught to respond with simple phrases (for example, "This is a cup" or "The ball is in the box")." [Watson et al., 1989]. Generally, behavioural approaches emphasise defining the conditions of positive and negative reinforcement, comprising 1) the cues and prompts to be given to the children, 2) the response the child is expected to show, and 3) the reinforcement to be given to the child after a correct response and the consequences for incorrect responses. Traditionally, in behavioural intervention little attention has been paid to how to use the forms in everyday communication.

## TEACHING SPONTANEOUS COMMUNICATION WITH TEACCH

The widely used spontaneous communication curriculum TEACCH was developed in North Carolina's TEACCH Division (Treatment and Education of Autistic and related Communication handicapped CHildren) and is influenced by both psycholinguistic and behavioural approaches [Watson et al., 1989]. Like behavioural approaches structure is emphasised, specific behaviours are targeted, conditions and consequences of eliciting the behaviour and defined, and behaviour is shaped through the use of cueing and prompting. Psycholinguistic approaches have influenced the categories of communicative behaviour which are used for assessment and defining teaching objectives, and the choice of objects which are developmentally appropriate for an individual student. Functionality (behavioural view) and pragmatics (psycholinguistic view) are the central issues in the TEACCH methodology. "More meanings for more purposes in more situations" are taught prior to teaching communication with more complex forms [Watson et al., 1989]. Naturalistic, less structured settings with naturalistic consequences are preferred to artificial settings. The TEACCH curriculum addresses a wide spectrum of communicative functions (request, get attention, reject or refuse, comment, given information, seek information, express feelings, social routine) and forms of communication (motoric, gestural, vocal, pictorial, written, sign, verbal). An example of a teaching session, which is supposed to encourage social interaction, addresses turn-taking behaviour between two children: two children and a teacher are sitting around a table, on the table an interesting toy that the children can operate with a marble. Turn-taking behaviour and handing over of the toy is encouraged in a 'playful' situation, under guidance by the teacher.

The mobile robot in the AURORA project is similarly using turn-taking, emphasising the dynamic aspects of interaction, but in a context where the children are free to move. We address a group of children where a central teaching objective is to encourage the child to ‘ask’ in order to get help, assistance, food etc. (see communicative functions mentioned above). Interactions with the robot are supposed to address e.g. the following therapeutic issues:

- Attention span during interaction with the robot, compared to the attention span children show in other contexts
- Robot-child ‘eye-contact’, i.e. the child making eye-contact with what is perceived as the robot’s ‘front’ (indicated by the robot’s preferred direction of movement and sensors located at the front end of the robot’s chassis)
- Pro-active behaviour towards the robot in order to elicit certain behaviour, and interaction
- Usage of behavioural and verbal cues in pro-active behaviour
- Turn-taking and imitative interaction games
- Increase in play and language skills at the expense of ritualistic behaviour

## STEREOTYPICAL AND REPETITIVE BEHAVIOUR

Most people with autism show aspects of rigidity and routine throughout their lives which becomes visible by a resistance to change the daily routine. They prefer a stable, ordered and secure environment. They are to varying degrees prone to stereotyped and repetitive movements like spinning or frequent touching of objects, rocking, headbanging, aligning objects in endless lines, obsessive collecting behaviour or attachments, verbal routines, and other repetitive behaviours. Aversive and punitive methods of treatments have largely become abandoned, while variations of the ‘theme’ of obsession (e.g. introducing restrictions and limitations to the execution of the behaviour) can prove successful [Howlin & Rutter, 1987]. It is not yet clear what causes this behaviour. In non-autistic people (and other animals) repetitive behaviour can be caused by sensory-deprivation. Turner [Turner, 1997] proposes that an executive dysfunction (inability to generate novel behaviour and ideas) can explain repetitive behaviour in individuals with autism. Alternatively, repetitive behaviour can be considered in terms of escape from (visual or auditory) over-stimulation which many people with autism suffer from. If the world appears chaotic, unpredictable and frightening and cannot be understood, then repetitive movements, in particular if they involve the own body (which can be pleasurable on the level of bodily experience), can have a calming effect in providing a safe and predictable environment. A robot, which can be easily programmed to perform repetitive and ‘predictable’ movements on the one hand, and to investigate variations of these movements on the other hand, might establish a link between the child and the world around it.

The success of the AURORA project will crucially depend on the careful design and investigation of a robot as a social mediator which can help children with autism to get used to basic styles of interaction which are typical of human beings. Thus, the robot will be introduced to the child with a fixed behaviour repertoire, but in interaction with the child it will slowly express a greater variety and encourage corresponding responses in the children. Ultimately, the child might get more and more used to interacting with an agent showing elements of unpredictable behaviour, and might ultimately be more and more confident and motivated to interact with humans who are joining the child-robot interaction ‘game’.

The project is also expected to give us insight in general issues of human-robot interaction that can be applied in different applications, for example with respect to how the interaction dynamics can be exploited in successful social interactions. Modelling a social agent need not require an explicit model of another agent; it can be implicitly ‘coded’ in the interaction dynamics that emerge in a concrete social interaction [Dautenhahn 1998a].

## THE SCATTERED WORLD OF AUTISM

A major problem of autistic people is to make sense of the world, figuring out what is going on around them. Most obvious is hereby their difficulty in understanding the social world, which has been discussed in the previous sections. However, a central problem which individuals with autism have with the social and non-social world likewise, described e.g. in [Grandin, 1995], [Varma, 1996], is their difficulty in perceiving objects in the world in a holistic sense, e.g. integrating different views of the same object or human. The world is perceived in unconnected parts, concepts of objects have to be laboriously synthesized, in order to create meaning, i.e. to order to recognize familiar faces and places. Spontaneous recognition is limited to parts, not the objects themselves or the 'whole'. This affects all aspects of daily life action and interaction, ranging from difficulties in interpreting pointing gestures (holding an object, such as a pen, up and saying "look at this" does not necessarily mean that the individual with autism looks at the pen, he can look at your hand, the background wall, or any other detail he perceives, even when looking in the direction where the object is located) to fundamental misunderstanding in daily routine dialogue (an individual with autism will spontaneously answer the question "Do you know your telephone number?" with "Yes", unaware of the 'larger picture' which identifies the question as a particular request to give information). Similarly, when learning a particular task, e.g. using a spoon to eat ice cream, children with autism are likely to have difficulty in generalising the task, i.e. using the spoon for other food items.

People with autism show us how much we usually rely on the power of generalization and abstraction which is necessary for understanding the world, whereas they may experience and recount experiences as a linear sequence of events: without reliably knowing that the person which I'm talking too is 'you' (since the person might have got a new haircut or is wearing a new hat) the 'meaning' of the social interaction is not clear. Humans use language (in social interaction) as an important means to form bonds, to get attached to other people, to identify correspondences between own experiences and experiences of other persons (sympathy, empathy). Getting engaged with another person requires identifying the other person in a holistic sense, only then the 'you', having a particular identity and 'biography' (e.g. experiences associated with the person) can be interpreted with respect to past, present and future experiences.

For researchers studying the scattered world perceived by people with autism can therefore yield significant insights into human perception, which under usual circumstances seems so amazingly efficient in perceiving parts as a whole, recognising patterns and similarities, recognizing objects even under circumstances where very little 'training examples' were available. This integrative, constructive, and holistic nature of human perception (let alone other aspects of human embodiment and cognition) is needed in successfully interacting with the social and non-social world. The AURORA project has to address to some extent these issues in the way human-robot interactions are designed, i.e. we cannot rely on the holistic nature of human perception which can usually compensate for a number of deficiencies in human-machine interfaces.

## STATE-OF-THE-ART OF THE AURORA PROJECT

More technical details and state-of-the-art of the AURORA project are described in [Werry & Dautenhahn, 1999]. Initial trials were promising with respect to how children interacted with the robot. A major encouragement was the fact that most of the children visibly enjoyed the interactions, supporting the approach to use the robot in a playful context. Current efforts are to enhance the types of behaviours that the robot can exhibit (including physical movements and speech recognition/synthesis), and to develop a vision system which can be used in order to enhance the sensory capabilities of the robot. The latter is necessary in order to allow the robot to access information on the location and movements of the child. The movements of the child are completely unrestricted and only limited by the fact that both robot and child are in the same room. The child can walk around, crawl on the floor, go down on its knees or can adopt any body posture in any orientation and distance to the robot. This poses serious demands on the flexibility and real-time constraints. A control architecture is under development which can integrate sensor processing and control of actuators in order to have the robot enacting a particular interactive story which the children can relate to, enjoy interacting with, and that leads to interactions which have a therapeutic effect.

The central goal of the AURORA project is to develop the robot as a teaching device in order to help individuals with autism to make contact with the social world of non-autistic people. One long-term challenge might be to investigate if it is feasible and desirable to develop a robot as a social buffer or mediator between the individual with autism and the world. In the same way as a seeing-eye dog's eyesight empowers blind people, a robot can be regarded as a means of social communication which can interact appropriately both with autistic and non-autistic people. Since autism is a minority handicap we cannot hope that non-autistic people learn how to communicate with autistic people. For people with autism, on the other hand, learning the rules of social communication and interaction in order to adapt to the world of the non-autistic represents a huge learning effort. Instead, integration of autistic and non-autistic groups of people could be enhanced by allocating the laborious task of entrusting 'correct' communication between autistic and non-autistic individuals to the robot, or other technological means of communication.

## (NON-) HUMANOID ROBOTS IN REHABILITATION

This section discusses application areas in which humanoid and non-humanoid robots can be advantageous. There are many exciting aspects to humanoid robots, for instance:

- building intelligent robots with humanoid body which can be used in contexts where a humanoid body is advantageous and desirable (e.g. for 'face-to-face' communication with a human).
- using them as a test-bed in order to study theories and models in developmental psychology and biology (see [Brooks et al., 1999]), e.g. studying hand-eye control, imitation.
- using them as interactive agents in contexts where humans expect and enjoy a 'natural' interface. In rehabilitation humanoids could replace traditionally used machines with 'machine-characters', e.g. in physiotherapy. For the user this could make interactions with the machine more enjoyable and interesting, and add a 'humane' aspect to the interactions, e.g. allowing communication via natural language, gestures, eye contact etc., letting the robot explain and express its actions via facial expressions, story-telling etc. Humanoid robots can make machines more interesting for people who have the tendency to anthropomorphise, e.g. people with physical or learning disabilities without specific impairments of imagination or social skills.

Non-humanoid robots are advantageous in the following contexts:

- application areas where the 'naturalness' of the humanoid interface does not outweigh the speed and efficiency of interaction and control. Humans are not able to communicate via directly networked brains. They have to interact via the often time-consuming and complex process of

- perception (speech understanding, recognition of body postures, gestures, facial expressions etc.)
- interpretation, ‘understanding’ of verbal and non-verbal cues
- appropriate and believable expression of a response, verbally and non-verbally

A more ‘artificial’ interface (e.g. controlling a robot by pressing buttons or keyboard input) can in certain contexts be faster, safer and more reliable and efficient. If a machine is only used for very few and clearly defined tasks then it might be faster to control it by pressing buttons, e.g. a vacuum cleaner.

- application areas where the humanoid appearance is misleading since the robot has only very limited capabilities. A humanoid robot is expected to behave humanoid, and if it does not the humanoid appearance can be misleading. E.g. dressing a vending machine like a humanoid does not seem to be promising.
- application areas where interaction partners do not accept the robot as a social partner. For example a household robot which has to deal with the family pets cannot hope that a dog accepts it as a human member of the pack. Dog perception is strongly guided by olfactory cues (in contrast to humans who are very ‘visual’ animals and can be easily fooled by a carefully designed look). Thus, as long as the humanoid does not smell like a human we cannot hope to fool a dog. If the robot’s task is therefore to look after the pets while the pet owners are on holiday, non-human types of communication can be more successful.
- machines which are interacting with people with autism where machine-type machines can be more interesting and less frightening than humanoid machines, see issues discussed in previous sections.

Thus, different application areas seem to require different designs of interactive robots, not only due to the functionality of the system but also due to the way humans expect and perceive the interactions. Analysis of such issues, which has been done in Human-Computer Interaction research with respect to interface design and software agents, similarly applies to robot-human interaction, see an example study in [Bumby & Dautenhahn, 1999]. As Reeves and Nass have shown [Reeves & Nass, 1996] humans tend to treat computers (and media in general) as people. This ‘media equation’ (media equal real life) is particularly relevant for robots whose physical appearance stresses ‘realism’. Robots can potentially be even more ‘persuasive’ than computers are [Fogg, 1999]. Robots are not yet part of our daily life in the same way as computers already are, but like computers they have the power to change our behaviour, attitudes, motivations, or worldview, and therefore belong to what is called in [Fogg, 1999] ‘persuasive technology’. Like computers, robots can serve as tools (providing users with new abilities), medium (conveying symbolic or sensory content), or social actors (invoking social responses from users). Issues of design, credibility [Tseng & Fogg, 1999], and ethics of persuasive technology [Berdichevsky & Neunschwander, 1999] therefore apply to robot technology, in particular to the new generation of interactive ‘social’ robotic systems.

## CONCLUSION

The paper discussed the questions whether agents, in particular robots, necessarily have to model human behaviour in order to achieve successful interactions. We focussed on the issue of predictability of behaviour, which is a great strength of robots but not of humans. In different application domains agents might be preferred which are human-like, or machine-like. Predictability of behaviour is an issue that we investigate in a project with humans whose social and imaginative skills are impaired, so we cannot strongly rely on anthropomorphism and empathic responses which usually can make a robot interesting in the eyes of the observer. Thus, this particular project forces us to be very explicit about properties like predictability and the particular way how children with autism perceive and interact with the world. The goals of the project AURORA are twofold: 1) helping children with autism in making steps to bond with the (social) world, 2) studying general issues of human-robot interface design with the human-in-the-loop, in particular a) the dynamics of perception-action with respect to both the robot and the human, b) the role of verbal and non-verbal communication in making interactions 'social', c) the process of adaptation, i.e. humans adapting to robots as social actors, and robots adapting to individual cognitive needs and requirements of humans as social actors.

The AURORA project does not intend to replace humans by robots. The ultimate goal is to help children with autism in making sense of the world, assisted by computer and robotic technology. The robot might become a 'friend' in a playful context, but emotional bonding with the robot is not our goal. Emotional bonding with humans or animals is far more fulfilling and reciprocal than bonding with a machine made of metal and silicon. After careful introduction, animals can play a very successful part in therapy. However, animals are not very suitable to be used as a teaching device on a routine basis since they cannot be controlled and instructed in the way a robot can, and need special care and treatment which requires substantially more effort than the maintenance of a robot.

To conclude, social robots as mediators for people with autism need not necessarily mimic human appearance. Being easily identifiable as a machine can in this context, where we cannot rely on anthropomorphism, be an advantage.

Generally, designing life-like robots which closely mimic animals or humans risks unnecessarily restricting and narrowing the apparent and actual functionality of the robot. To give an example: A child who bonds with the family's robot dog whose task is to protect the home, might be shocked to see her mother using the dog as a 'mobile' phone. Here, a multi-functional 'machine-like' robot performing exactly the same tasks would appear more consistent, less disturbing, and ultimately more believable and credible to humans interacting with it.

## ACKNOWLEDGEMENTS

The project AURORA is conducted in collaboration with Radlett Lodge School, a school of the UK National Autistic Society (NAS). The project is supported by The Engineering and Physical Sciences Research Council (EPSRC). We are grateful to Takashi Gomi and his team at *Applied AI Systems Inc.* for continuing support and encouragement. Some issues addressed in this paper are investigated by Iain Werry as part of his PhD thesis supervised by the author.

## REFERENCES:

AURORA: http://www.cyber.rdg.ac.uk/people/kd/WWW/aurora.html, last referenced on 1st of June, 1999

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985) Does the autistic child have a 'theory of mind'? *Cognition*, 21, pp. 37-46.

Berdichevsky, D. & Neunschwander, E. (1999) Toward an ethics of persuasive technology. *Communications of the ACM*, vol 42(5), May 1999, pp. 51-58.
Bettelheim, B. (1967) *The empty fortress: infantile autism and the birth of the self*. New York: Free Press.
Billard, B. & Dautenhahn, K. (1998): Grounding communication in autonomous robots: an experimental study. *Robotics and Autonomous Systems*, Special Issue on "Scientific methods in mobile robotics", M. Recce and U. Nehmzow (Eds.), no. 24, vols. 1-2, pp. 71-81.
Breazeal, C. (1998) A Motivational System for Regulating Human-Robot Interaction. In *Proc. of AAAI 98*, Madison, WI.
Breazeal, C. & Foerst, A. (1999) Schmoozing with robots: Exploring the boundary of the original wireless network. Proc. *The Third International Cognitive Technology Conference* (CT99), August, San Francisco.
Brooks, R.A., Breazeal, C., Marjonavic, M., Scassellati, B., & Williamson, M. M. (1999) The Cog project: building a humanoid robot. In: C. L. Nehaniv (Ed.) *Computation for Metaphors, Analogy, and Agents*, vol. 1562, Lecture Notes in AI, Springer Verlag, pp. 52-87
Bumby, K. & Dautenhahn, K. (1999) Investigating Children's Attitudes Towards Robots: A Case Study. Proc. *The Third International Cognitive Technology Conference* (CT99), August, San Francisco.
Byrne, R. W. (1995) *The thinking ape, evolutionary origins of intelligence*. Oxford University Press.
Byrne, R. W. & Whiten, A. (Eds.) (1988) *Machiavellian intelligence : social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press
Cypher A. (1993) *Watch what I do: Programming by demonstration*. MIT Press.
Dautenhahn, K. (1997) I could be you - the phenomenological dimension of social understanding. *Cybernetics and Systems Journal*, 28(5), pp. 417-453.
Dautenhahn, K. (1997b) The Role of Interactive Conceptions of Intelligence and Life in Cognitive Technology. *Proc. Second International Cognitive Technology Conference*, (Eds.) Marsh, J. P., Chrystopher L. Nehaniv and Barbara Gorayska, CT'97, 25-28 August 1997, Aizu, Japan, IEEE Computer Society Press, pp. 33-43
Dautenhahn, K. (1998a): Grounding Agent Sociality: The Social World is its Own Best Model. *Proceedings of 14th European Meeting on Cybernetics and Systems Research*, Editor: Robert Trappl, pp. 779-784.
Dautenhahn, K. (1998b): The Art of Designing Socially Intelligent Agents - Science, Fiction, and the Human in the Loop. Special Issue "Socially Intelligent Agents", *Applied Artificial Intelligence Journal*, vol 12, 7-8, October-December, pp. 573-617.
Dennett, D.C. (1987). *The Intentional Stance*, Cambridge MA. MIT Press
Dryer, D. C. (1999) Getting personal with computers: how to design personalities for agents. *Applied Artificial Intelligence Journal*, Special Issue on "Socially Intelligent Agents", vol 13 (2).
Eddy, T. J., Gallup, G. G., & Povinelli, D. J. (1993) Attribution of cognitive states to animals: anthropomorphism in comparitive perspective. *Journal of Social Issues*, vol. 49(1), pp. 87-101.
Fogg, B. J. (1999) Persuasive technologies. *Communications of the ACM*, vol. 42(5), May 1999, pp. 27-29.
Foner, L. N. (1997) Entertaining agents: a sociological case study. *Proc. of the First International Conference on Autonomous Agents*, Marina del Rey, CA, USA, February 5-8, (Ed.) W. Lewis Johnson, pp. 122-129
Galef, B. G. & Heyes, C. M. (Eds) (1996) *Social learning in animals: the roots of culture*. Academic Press, New York.

Gaussier, P., Moga, S., Banquet, J.P., & Quoy, M. (1998) From Perception-Action Loops to Imitation Processes: A Bottom-up Approach of Learning by Imitation. *Applied Artificial Intelligence Journal*, Special Issue on "Socially Intelligent Agents", vol 12 (7-8), pp. 701-729.

Grandin, T. (1995) *Thinking in pictures.* Doubleday Publisher.

Grandin, T. & Scariano, M. M. (1996) *Emergence: labeled autistic*. Warner Books.

Hayes, G. & Demiris, J. (1994) A robot controller using learning by imitation. *Proc. International Symposium on Intelligent Robotic Systems*, Grenoble, pp. 198-204.

Hendriks-Jansen, H. (1997) The Epistemology of Autism: Making a Case for an Embodied, Dynamic, and Historic Explanation. *Cybernetics and Systems Journal*, vol 28(5), pp. 359-416.

Howlin, P. & Rutter, M. (1987) *Treatment of autistic children.* John Wiley and Sons.

Lovaas, O. I. (1977) *The autistic child*. New York: Irvington.

Loyall, B. & Bates, J. (1997) Personality-rich believable agents that use language. *Proc. of the First International Conference on Autonomous Agents*, (Ed.) W. Lewis Johnson, pp. 106-113.

Meltzoff, A. N. (1996) The human infant as imitative generalist: a 20-year progress report on infant imitation with implications for comparative psychology. In B. G. Galef, C. M. Heyes (Eds.) *Social learning in animals: the roots of culture*, Academic Press, pp. 347-370.

Moravec, H. (1988) *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

Nehaniv, C. L. (1999) The Second Person - Meaning and Metaphors. In: C. L. Nehaniv (Ed.) *Computation for Metaphors, Analogy, and Agents*, vol. 1562, Lecture Notes in AI, Springer Verlag,, pp. 380-388.

Reeves, B. & Nass, C. (1996) *The Media Equation*. Cambridge University Press.

Russell, J. (Ed.) (1997) *Autism as an Executive Disorder*, Oxford University Press.

Shneiderman, B. (1992) *Designing the user interface. Strategies for effective human-computer interaction.* Addison-Wesley Publishing Company.

Trappl, R. & Petta, P. (1997) *Creating personalities for synthetic actors*. Springer.

Tseng, S. & Fogg, B. J. (1999) Credibility and computing technology. *Communications of the ACM*, vol. 42(5), May 1999, pp. 39-44.

Watson, L. R., Lord, C., Schaffer, B., & Schopler, E. (1989) *Teaching spontaneous communication to autistic and developmentally handicapped children*. Irvington Publishers, Inc., New York.

Turner, M. (1997) Towards an executive dysfunction account of repetitive behaviour in autism. In Russell, J. (Ed.) *Autism as an Executive Disorder*, Oxford University Press.

Varma, V. P. (Ed.) (1996) *The Inner Life of Children with Special Needs*. Whurr Publishers Ltd, London.

Watt, S. N. K. (1997) *Seing things as people – anthropomorphism and common sense psychology*. PhD Thesis, KMI-TR-61, Knowledge Media Institute, The Open University, UK.

Werry, I. & Dautenhahn, K. (1999) Applying robot technology to the rehabilitation of autistic children. *Proc. Symposium on Intelligent Robotics Systems* (*SIRS'99*), 20-23 July 1999, Coimbra, Portugal.