

Socially Intelligent Agents and The Primate Social Brain – Towards a Science of Social Minds

Kerstin Dautenhahn

Adaptive Systems Research Group
Faculty of Engineering & Information Sciences
University of Hertfordshire
Hatfield, Herts AL10 9AB, United Kingdom
K.Dautenhahn@herts.ac.uk
<http://homepages.feis.herts.ac.uk/~comqkd>

Abstract

This article puts research on socially intelligent agents (SIA) in the broader context of how humans (and other primates) perceive and interact with the social world. Phylogenetic (evolutionary) and ontogenetic (developmental) issues are discussed with respect to the social origin of primate and human intelligence and human culture. Implications for designing artifacts and for the evolvability of human societies are outlined. A theory of empathy is presented that is based on current research on the primate social brain. Research projects that investigate some of these issues are reviewed. I argue that Socially Intelligent Agents (SIA) research, although strongly linked to software and robotic engineering, goes beyond a software engineering paradigm: it can potentially serve as a paradigm for a science of social minds. A systematic and experimental investigation of human social minds and the way humans perceive the social world can result in truly social artifacts, that are integrated in human society.

“Once there lived a robot called Rob, he was made by a mad professor called Brain-Box on Jan the 3rd 2,000,000 in Germany. Everybody thought it was brilliant and on Jan 30th the same year it ran away and it ran away to England and terrorised England people looked to see who or what had done it. The robot had killed someone and taken her brain they got the police out and they looked into it the police didn’t know who or what had done it. One day someone called Tod saw Rob and called the police to come and get Rob but when the police cam Rob had gone. Tod got a fine for supposedly lying to the police. The next year the police had still not found Rob. Sometime in the last year Rob had fallen off something and he was found in pieces in a junk yard. Some people phoned the police the police came out and he had gone. The village dug a massive pit to try and catch Rob. Now Rob was as high as six houses he was eating a lot of junk metal. One day rob was walking along and he fell in the pit, the people found him in the pit and he was killed.” (Christopher, 9 years old, (BD99)).

Introduction: Socially Intelligent Agents - The Human in the Loop

As Reeves and Nass have shown (RN96) humans tend to treat computers (and media in general) as people. I believe that this ‘media equation’ (media equals real life) is particularly relevant for socially intelligent agents (SIA’s) research with the ‘human in the loop’, namely studying the relationship between socially intelligent agents and humans as designers, users, observers, assistants, collaborators, competitors, customers, or friends. In order to acknowledge the ‘human in the loop’ I suggested in (Dau98) a list of design guidelines for SIA technology, identifying the following roles of humans and suggesting that a balanced design of socially intelligent agents need to address these roles: Humans are embodied agents, humans are active agents, humans are individuals, humans are social beings, humans are storytellers, humans are animators, humans are autobiographic agents, humans are observers. Not unsurprisingly, SIA research is more than other agent research strongly inspired and motivated by findings outside software engineering and computer science, in particular the humanities, social sciences, and natural sciences. As such, SIA research is different from the field of agent-based computing research that views agents primarily as a software engineering paradigm (cf. (JSW98), (WJK00)). In contrast, SIA research benefits from viewing agents in the larger picture of autonomous agents as defined by Franklin and Graesser ((FG97)): “An autonomous agent is a system situated within and a part of the environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to affect what it senses in the future”. This definition is very attractive since it applies easily both to natural and artificial systems. Franklin and Graesser (FG97) also propose a natural kinds taxonomy of agents, distinguishing between biological, robotic and computational agents as instances of the class of autonomous

agents. In a different paper Stan Franklin argues for the study of autonomous agents as embodied artificial intelligence (Fra97)¹. I believe that this was an important step, namely viewing (autonomous) agents as vehicles and subjects for the investigation of artificial (and human) intelligence, and in this paper I like to argue that socially intelligent agents can similarly serve as tools and vehicles for the study of artificial (and human) social minds.

Socially Intelligent Agents research is concerned with agents and human-agent interactions whereby the agents show aspects of human-style social intelligence.

Socially intelligent agents are biological or artificial agents that show elements of (human-style) social intelligence. The term *artificial social intelligence* refers then to an instantiation of human-style social intelligence in artificial agents. (Dau98)

Social intelligence can be natural (humans) or artificial (computational or robotic agents), but within the context of human-style social interaction and behaviour. Please note that social intelligence in this sense does not make claims on how intelligent the agent needs to be: often simple strategies can be socially very effective!

Although SIA research is primarily interested in human-style behaviour and interactions, human social intelligence has a history, both evolutionary as well as developmental. Below I therefore discuss a few findings from primatology and developmental psychology and their implications for SIA research.

The social, ethical, cultural, as well as cognitive implications of SIA technology are important issues to consider. Even on the level of the individual human, interactions with SIA's can influence a human beings attitudes, behaviour and minds, and in this way empower as well as manipulate humans (see discussion in (ND00)). In (Fog99) B.J. Fogg discusses computers as *persuasive technologies*. In contrast to other non-persuasive technologies, "persuasive computing technology is a computing system, device, or application intentionally designed to change a person's attitudes or behaviour in a predetermined way." Furthermore,

¹Embodiment is here considered as 'embodied in the situated sense of being autonomous agents structurally coupled with their environment'. In this sense software agents can be as embodied as robotic and biological agents. Tom Quick developed a definition of embodiment based on structural coupling and mutual perturbation between an agent and its environment, a definition that applies to different kinds of systems, including autonomous agents, and which allows to measure different degrees of embodiment quantitatively (QDNR99b), (QDNR99a), (QDNR99c).

Fogg calls the study of planned persuasive effects of computer technologies *captology*. Figure 1 shows the functional triad of computer persuasion. Following Fogg's terminology SIA's might fall under the category of 'social actors', where agents can adopt animate characteristics, play animate roles, and follow social dynamics for the purpose of creating relationships with humans and invoke social responses. In this sense, SIA's are persuasive technologies and therefore issues of design, credibility (TF99), and ethics of persuasive technology (BN99) also apply to SIA's technology, in particular to the new generation of highly interactive 'social' software and robotic agents, many of them described in this volume and elsewhere.²

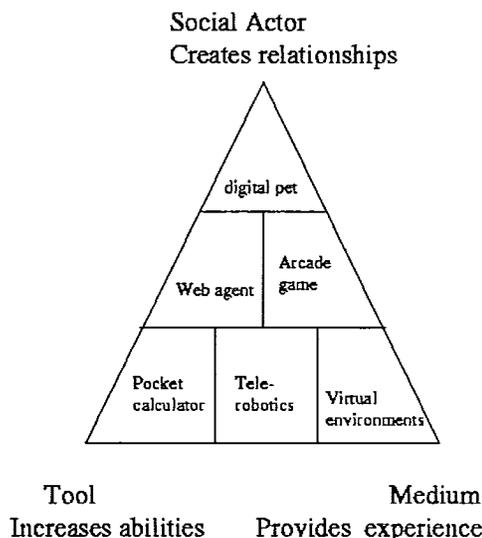


Figure 1: The functional triad of computer persuasion, redrawn from (Fog99).

The Life-Like Agents Hypothesis

As discussed in the previous section, SIA's are often designed to 'imitate' life. Based on what I called previously the 'Life-Like Agents Hypothesis' this approach can be characterised as follows (Dau99c):

²Examples of collections of articles on SIA research are: K.Dautenhahn, C. Numaoka (guest editors): "Socially Intelligent Agents", Special Issues of Applied Artificial Intelligence, Vol 12 (7-8), 1998, and Vol 13 (3) 1999, K.Dautenhahn (2000) Human Cognition and Social Agent Technology, John Benjamins Publishing Company, B. Edmonds and K. Dautenhahn (guest editors): Social Intelligence, special issue of Computational and Mathematical Organisation Theory, 5(3), 1999, K. Dautenhahn (guest editor): Simulation Models of Social Agents, special issue of Adaptive Behaviour, 7(3-4), 1999

“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, and 3) Life-like agents can serve as models for the scientific investigation of animal behaviour and animal minds”.

Argument (3) is certainly valid and need not be discussed here. However, arguments (1) and (2) are not as straightforward as they seem. Designing life-like agents that closely mimic human appearance or behaviour can unnecessarily restrict and narrow the apparent and actual functionality of an agent. Similarly, imagine mobile phones were designed so that they had the shape of old-fashioned dial-operated telephones. It then could be disturbing or at least puzzling for people to find out that the mobile phone might have more functionalities (e.g. sending and receiving email, browsing the Web etc.) than the original model. Thus, ‘new’ designs not imitating any other previously existing object might better suit a piece of technology that is combining functionalities in a novel way or has new functionalities. A social interface agent (e.g. in an e-commerce context) presented with humanoid appearance and behaviour might have the advantage of evoking an initial feeling of ‘familiarity’ in a human customer, but 1) human customer’s are then likely to expect the agent to show other human characteristics and functionalities, human knowledge, personality and other characteristics of humans in general (including that it understands jokes and possesses common sense knowledge), and sales agents in particular, and 2) new or different functionalities that the real agent does not possess need to be integrated in a plausible way in the agent’s behaviour, without breaking the suspense of disbelief (Mat97) (see (ND00) for further discussion of these issues).

Attitudes towards Socially Intelligent Agents: Anthropomorphism and Behaviour Reading

According to the *Social Intelligence Hypothesis*, discussed in more detail below, the evolution of primate intelligence is linked with an increase of the complexity of primate social life ((BE98), (WE97)). The argument suggests that during the evolution of human intelligence a transfer took place from social to non-social intelligence so that hominid primates could transfer their expertise from the social to the non-social domain (see review in (Gig97)). An interesting aspect of this kind of transfer is given by Mithen (Mit96), who explains the evolution of anthropomorphic thinking with an accessibility between the domains of social intelligence and natural history intelligence so that “people could be thought of as animals, and animals could be thought of as people”, (Mithen 1996, p. 224).

The attribution of human motivation, characteristics, or behaviour to inanimate objects, animals or natural phenomena is usually called anthropomorphism (see The American Heritage©Dictionary of the English Language). Anthropomorphism is often dismissed as a curiosity or unscientific phenomenon and only relatively few scientific work outside philosophy have experimentally addressed the issues of how and why people tend to adopt an *intentional stance* (Den71), (Den87), namely treating non-human objects and animals as intentional objects (what seems to be based on the human mindreading or social competence system, see discussion below). It is often suggested that physical likeness, familiarity, phylogeny and/or cultural stereotypes are important factors. Well known is the study by Eddy et al. (EGP93) who investigated peoples tendency to anthropomorphise animals (see summary in (Wat97)). The study identified two primary mechanisms why people attribute similar experiences or cognitive abilities to animals, based on 1) the degree of physical similarity, and 2) the degree of an existing attachment bond (familiarity). Dogs and cats are more familiar to most people than frogs, primates are physically (and behaviourally) similar to humans.

This study seems to support the above mentioned *Life-Like Agents Hypothesis*, namely that humanoid agents that look like humans should be more believable and successful as social interaction partners for humans than non-humanoid agents (assuming that humans mostly enjoy interacting with other humans). However, other evidence suggests that not physical similarity, but *behaviour in context* matters. Mitchell and Hamm (MH97) provided undergraduate students with narratives depicting different mammalian agents

(including humans) showing behaviour that suggested jealousy or deception. The students were then asked to answer questions on particular psychological characterisations of the agents. The narratives varied according to species, context in which an agent's behaviour occurred, and the degree of emphasis that the narrative was about a particular species of animal (or human). The behaviour was constant in all narratives. Mitchell and Hamm found that variations in context influenced the psychological characterisations, but variations in species and emphasis did not, i.e. the psychological characterisations of all species were almost always similar: "Nonscientists (and some scientists as well) apparently use a mammal's behaviour-in-context (whether human or not) as evidence of its psychological nature, regardless of the mammal's physical similarity, familiarity, or phylogenetic closeness to humans, or the mammal's cultural stereotype; psychological terms are not used specifically for humans, but rather are depictive of behaviour-in-context". Interestingly, the notion that behaviour matters more than appearance in ascribing intentionality is supported by an experimental study published in 1944 (HS44) that convincingly demonstrates the effects of the 'intentional stance'. Here, human subjects created elaborate narratives about intentional agents when asked to describe movements of moving geometric shapes shown in a silent film. Other studies along research done by Mitchell and Hamm and Heider and Simmel could confirm whether this also applies to non-mammalian animals. A particular challenge would be to include computational and robotic agents in such studies. I suggest that behaviour-reading might apply also to inanimate objects such as robots. Every robotics researcher who has ever given a demonstration of autonomous mobile robots to a general audience can confirm how readily humans view robots as people, cf. (Bra84), (BD99). The importance of behaviour expression in agent building has been recognised e.g. by Phoebe Senegers (Sen98), (Sen00). Her argument is that 'doing the right thing' (the classical approach of AI approaches to agent control) needs to be complemented by paying attention to 'doing the thing right', in particular creating believable transitions between agent behaviours.

Attitudes Towards Agents: A Case Study with Robots

In (BD99) Kate Bumby and Kerstin Dautenhahn investigated children's attitudes towards robots, a brief summary is given here. We were interested to find out how children interact and describe robots. Thirty eight children (ages seven to eleven, 21 males, 17 females, BC1 socioeconomic category) were studied at

St. Margarets Junior School in Durham, UK. A number of working hypotheses were addressed with respect to how the children portrayed robots. In three studies the children were asked a) to draw a picture of a robot, b) to write a story about the robot they had drawn. These studies were observational. The third study had the format of an informal, guided and filmed interview while the children were in a group interacting with two mobile robots (see figure 2) that were running in an environment with a lightsource. The robots were simple behaviour-based vehicles (Bra84).

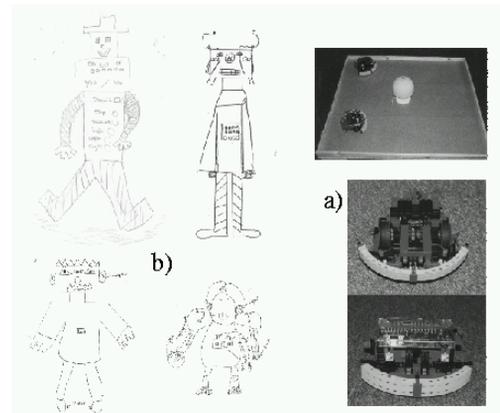


Figure 2: a) The experimental set up and the two autonomous, mobile *fischertechnik* robots, b) Drawings of 8-year and 9-year olds, (BD99).

Results of study a (pictures) show e.g. that the children tend to give the robot humanoid faces. Figure 2 shows examples of a variety of drawings by 8/9-year olds, portraying robots. In study b (stories), one result was that the children tend to put the robots in familiar settings, doing familiar tasks. The robots were significantly often put in a social context. Study c (interview) showed a clear tendency to anthropomorphise the robots, e.g. "I don't think it likes the light.". The children also often talked to the robots as if they were animals or small children. Other findings of this case study, e.g. with respect to attribution of gender or violence is reported in more detail in (BD99). This single case study cannot answer the question of how children in that age range in general think about robots, but the results give some indication that confirm findings along the lines of studies with computers (RN96).

Societies of Social Animals³

Swarm Intelligence: Social Insects Don't have Friends

The term 'societies' is generally applied both to human and other animal societies, including social insects. Social insects (e.g. termites, bees, ants) are very well studied and two important theoretical concepts are used to understand coordination in social insect societies, namely *self-organisation* and *stigmergy*. Recently, models of *swarm intelligence* and their applications to problems like combinatorial optimisation and routing in communications networks have been studied extensively (BDT99). The concept *stigmergy* describes a class of mechanisms mediating animal-animal interactions, based on the description of insect behaviour as stimulus-response (S-R) sequences (even for solitary species). Stigmergy is based on indirect communication, communication via the environment, and an example of *collective behaviour*.

Primate Intelligence: Getting to Know Each Other

In primate societies, and different from members of social insect societies, an individual is not only socially situated (being part of and surrounded by a social environment) but also socially embedded (ED98) which means that the agent needs to pay attention to other agents and their interactions individually. Particularly human primates are specialised in predicting, manipulating and dealing with highly complex social dynamics (involving direct relationships as well as third-party relationships); they possess language as an effective means of preserving group coherence, 'social grooming' (Dun93), (BD97) and communicate about themselves and others in terms of stories (Dau99b). Humans are not only dealing with very complex relationships but seem to have mental 'models' of themselves, others and the social world (cf. (Whi91), (BC95)). Humans, different from social insects live in *individualised societies* (as do some other species of birds and mammals). An increasingly complex social field and an increasing need to effectively communicate with each other were likely to have been among the important constraints in the evolution of human minds.

Minds are certainly attributed to members of *Homo sapiens* (and as some evidence suggests several other hominid species might have existed with 'minds'), but other candidates exist among mammals (e.g. non-human apes, dolphins, elephants) and birds (e.g. parrots and members of the crow family). Interestingly, species which we describe as possessing a 'mind' are

³This section is based on (Dau00b).

all highly social. Even the 'solitary' life style of *Pongo pygmaeus* or orangutans, (who nevertheless seem to be highly social in their ability to recognise and interact with each other) is rather a secondary adaptation to a particular environment which demands a spatially 'distributed' social organisation. The *Social Intelligence Hypothesis* suggests that primate intelligence primarily evolved in adaptation to social complexity, i.e. in order to interpret, predict and manipulate conspecifics (see overview in (BE98), (WE97)). Thus, there are two important aspects to human sociality: it served as an evolutionary constraint which led to an increase of brain size in primates, this in return led to an increased capacity to further develop social complexity.

Although it is still unknown why hominids needed or chose to live in social groups, this *feedback principle* soon led to the development of highly sophisticated levels of organisation and control in human societies. In (Rus93) four levels of primate social organisation are discussed which might serve as models for the evolution of primate societies: a) the 'shrew'-type pre-primates: solitary, many offspring, insectivores, e.g. *Purgatorius*, a 70-million-year-old fossil, b) the 'mouse-lemur'-type primates: bush-living, nocturnal, strong mother-daughter bonding (stable matriline), social learning (offspring learns from mother), solitary males and social groups of mothers and daughters, e.g. the 50-million-year-old fossil *Shoshonius cooperi*, c) the '*Lemur catta*'-type diurnal lemurs: appearing about 54 million years ago, social groups (troops), dominant females, submissive males, stable matriline, occasionally consort bonds between single male and female, e.g. Adapidae, d) the 'chimpanzee'-type lemur-ape: appearing about 24 million years ago, groups of dominant males and submissive females, stable families of mothers and their offspring, male power coalitions, e.g. *Dryopithecus*. The social organisation of recent species of apes shows variations of this pattern: of harem-structures (gorilla), solitary lifestyle (orangutan). Such stages of social organisation can be related to behavioural as well as cognitive capacities of primates.

The terms 'theory of mind' and *mindreading* are usually used in order to discuss whether an animal is able to reflect on its own mental states (e.g. desires, intentions and beliefs) and those of others. Researchers have studied whether humans might have particularly specialised in a theory of mind (PW78), (PP95a). However, as Richard Byrne pointed out (Byr97), the *Social Intelligence Hypothesis* might account for the evolution of primate intelligence, but not for the specific human kind of intelligence. Here, narrative psychology and studies on the development of autobiographic memory

and a 'self' might offer an explanation: evidence suggests that 'stories' are the most efficient and natural human way to communicate, in particular to communicate about others (Bru91). Dennett (Den89) even regards the 'self' as a 'centre of narrative gravity'. Narrativity, the capacity to communicate in terms of stories is therefore regarded an efficient means to communicate social matters, and the origin of narratives might therefore have been a crucial milestone in the evolution of primate social intelligence (RM95). The *Narrative Intelligence Hypothesis* (Dau99b) proposes that the evolutionary origin of communicating in stories was correlated with increasing social dynamics among our human ancestors (see figure 3), in particular the necessity to communicate about *third-party relationships* (which in humans reaches the highest degree of sophistication among all apes, cf. gossip).

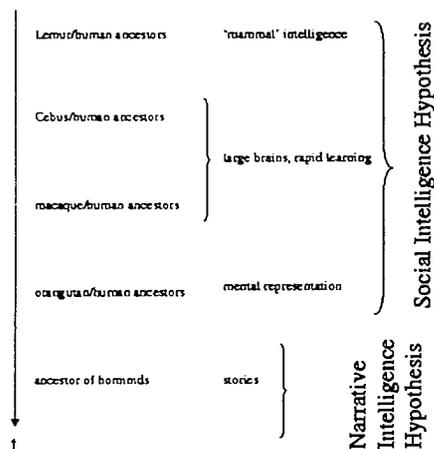


Figure 3: Evolution of the Social Mind, modified from (Byr97), see text for explanation.

A number of current research is devoted to building narrative software, virtual or physical environments (e.g. (GC97), (MMP99), (BC00), (BD99), (MAD+00), (BBA+00)). Supporting human narrative intelligence is expected to impact human minds and our notions of sociality and what we call our 'selves'. In parallel, investigations into autonomous story-telling agents can result in agents (robotic or software) with genuine narrative minds, being able to tell us interesting stories, listen to and understand our stories, and make us laugh. A first attempt to a bottom-up approach to narrative intelligence for autonomous agents is described in (CD00), (DC00). The kind of stories these agents might tell us will be shaped by the social field and the cultural environment of human societies

in which these agents 'grow up'. Thus, it's up to us whether the stories of the future will be nightmares, fairy-tales, comedies or adventures.

"Once they were a robot, that lived in the country. He came to this cliff, he thought it was too steep for him to go down so he went down the steps. He went onto the beach oh yeah the robot's name is Shaped and Shaped went to play in the sea and after a few minutes he fell to the ground and got washed into shore because the water has gone into his body." (Becky - 8 years old, (BD99)).

Primate Culture: We are not alone

The terms *anonymous* and *individualised* societies are used in biology in order to describe two different types of social organisation. Social insects are the most prominent example of anonymous (eusocial) societies where group members do not recognise each other as individuals but rather as group members⁴. We do not observe bees or termites searching for missing members of their colony. Although individuals adopt specific roles in a colony they do not show individuality or 'personality'.

The situation is quite different in individualised societies which primate societies belong among. Here we find complex recognition mechanisms of kin and group members. This gives rise to complex kinds of social interaction and the development of various forms of social relationships and networks. On the behavioural level long-lasting social bonding, attachment, alliances, dynamic (not genetically determined) hierarchies, social learning, development of traditions etc. are visible signs of individualised societies. In humans the evolution of language, culture and an elaborate cognitive system of mindreading and empathy are characteristics of human social intelligence in individualised societies (Dau97). As a consequence of the latter, humans are not only paying attention to other agents and their interactions individually (interactions between distinct personalities), but they use their mental capacities to reason about other agents and social interactions. It is at present unclear to what extent the social intelligence of members of other animal species, in particular very social species like elephants, Grey parrots, non-human apes and cetaceans, is similar to or different from our own. Similarly, the issues of cultural and 'memetic' evolution is highly controversial. The concept of memes, first introduced by Dawkins (Daw76) comprises ideas, fashions, skills and other components of *human* culture. Human culture and the memetic transmission of knowledge, ideas

⁴Note that African naked mole-rats, mammals, show a eusocial organisation similar to social insects (SJA91). Thus, the eusocial form of organisation has evolved independently in different taxa of animals.

and skills is often regarded unique to human societies. According to Donald's discussion of the evolution of culture and cognition (Don93) modern humans have three systems of memory organisation (mimetic skill, language and external symbols) not available to our primate relatives, and these 'inventive capacities' result in languages, gestures, social rituals, images etc. According to Tomasello et al. (TKR93) cultural learning is a uniquely human form of social learning. Cultural learning requires three social-cognitive processes which emerge in human ontogeny: imitative learning, instructed learning (teaching) and collaborative learning. Similarly, Blackmore (Bla99) argues that only sophisticated forms of imitation which are characteristic of humans but not non-human primates, were a necessary prerequisite for memetic replication which leads to human culture.

Others argue that culture as such is unlikely to be a feature unique to human societies and that the acquisition of novel behaviours in 'proto-cultures' can be observed in animals. To give an example: traditions have been observed among troops of Japanese macaque monkeys (Huf96): Japanese macaques or *Macaca fuscata* show several examples of the acquisition of innovative cultural behaviours, e.g. sweet potato washing and wheat-washing was invented in 1953 by a young female and subsequently spreading to older kin, siblings, and playmates, eventually to other members of the troop. Other observed cultural behaviours are fish eating (as many newly acquired food sources initially spreading from peripheral males to adult females, then from older to younger individuals), and stone handling or stone play (initially spread only laterally among individuals of the same age). Subsequently all these behaviours were passed down from older to younger individuals in successive generations (*tradition phase*). These examples clearly show the influence of social networks on the *transmission phase* of novel behaviour: the nature of the behaviour and social networks determine how the behaviours are initially transmitted, depending on who is likely to be together in a certain context and therefore is exposed to the novel behaviour. Innovative behaviours of the kind described here have been independently observed at different sites. Various factors have been discussed which influence cultural transmission: environmental factors, gender, and age, and other social and biological life history variables. For example, unlike potato or wheat washing, stone handling declines when individuals mature.

With respect to cultural transmission in non-human apes, recent evaluations of long-term field studies of chimpanzees or *Pan troglodytes* give compelling evidence for cultural behavioural variants (traditions) in

different chimpanzee communities, data which cannot be explained by ecological differences of the habitats, and comprising dozens of different behaviours including tool usage, grooming and courtship behaviours (WGM⁺99). Possibly the kind of *mechanisms* that are necessary for and support culture (e.g. cognitive mechanisms, language, imitation) might be different in different animal species. As Frans del Waal concludes (dW99): "The 'culture' label befits any species, such as the chimpanzee, in which one community can readily be distinguished from another by its unique suite of behavioural characteristics. Biologically speaking, humans have never been alone - now the same can be said of culture."

The striking similarity of cultural transmission of novel behaviour exhibited by Japanese macaque monkeys and chimpanzees and what we call human culture questions the uniqueness of human societies. Note that this behaviour is observed in monkeys, which do not show complex forms of social learning like imitation, and do not seem to possess higher-level 'cognitive' capacities necessary for complex social forms of 'primate politics' shown by non-human apes and humans. However, many non-human primates are very good social learners (widely using non-imitative forms of social learning, e.g. stimulus enhancement or social facilitation). Reader and Laland (RL99) therefore argue that the meme concept can and should also be applied to cultural transmission among non-human animals. Animal societies can appear in various forms. Human societies, human culture and human minds reflect in many ways their evolutionary origin in animal societies, animal culture and animal minds. Considering human culture in an evolutionary context, linking it to precursors in non-human primate societies might help a better understanding of human culture.

Implications for Evolvability of Human Societies

According to Kirschner and Gerhart evolvability can be defined as "the capacity to generate heritable phenotypic variation" (KG98). As outlined above, in primate non-human societies we already observe precursors of human culture (e.g. social learning, traditions). For reasons still under dispute our human ancestors were required to deal with increasingly complex social dynamics. Mental capacities evolved which allowed the evolution of increasingly complex mechanisms of social control, which in return increased the complexity of primate societies.

Based on what the previous sections discussed about primate societies and culture, the following requirements for mental capacities and social skills which fa-

cilitated the evolution of *primate culture* are proposed:

- Individualised societies: The capacity to identify and recognise individual group members.
- Social Networks: the capacity to establish, maintain, remember and utilise social networks. Three basic elements⁵ in the primate social field are:
 - Forming direct (one-to-one) relationships with group members
 - Identifying third-party relationships (relationships among other group members)
 - Recognition of conspecifics as members in a group hierarchy/social structure (e.g. structures of kinship, allies, dominance hierarchies, etc.)
- Efficient mechanisms of social bonding, either via physical grooming (in on-human primate societies) or via language and communication in narratives as efficient ways of ‘social grooming’, important for maintaining the coherence of social groups at different levels of social organisation.
- Social learning: the capacity to use others as ‘social tools’ (as explained in (Dau95)), via social learning mechanisms with varying degrees of what the animals learn from each other (cf. social facilitation versus imitation)

Thus, culture and other characteristics of human societies cannot be separated from specific environmental (including social) constraints and mental capacities which evolved as adaptations for dealing with such constraints. Specific adaptations then turned out to be prerequisites in the evolution of more sophisticated forms of primate societies and culture. Although new forms of media seem to substantially expand the social life of humans, even today the same mental capacities which were involved in the evolution of the human social animal now pose cognitive limits on the complexity and number of social encounters. Our primate social brain has a limit on the number of individuals who we can maintain direct social relationships with, namely relationships based on direct social knowledge (around 150), correlated with the relative size of the human neocortex (Dun93), (BD97). This figure can be identified consistently in various ancient and present human cultures. This number is significantly larger, namely more than double that observed in any population of non-human primates. Unless drastic (technological) enhancements of human cognitive capacities are invented, this number could only be exceeded by

⁵This is not supposed to be an exhaustive list.

inventing new, more efficient ways of “social grooming” (exceeding the communicative capacities of language). Another interesting issue discussed by Dunbar (Dun93), (BD97) is that language is 2.8 times more efficient as a mechanism of social bonding in comparison to physical grooming. The suggestion is therefore that human conversational group sizes should be limited to about 3.8 (which means one speaker and 2-3 listeners). Data on small group sizes confirm this hypothesis.

I showed above that biological evolution led to two distinctively different forms of social organisation in animal societies (anonymous and individualised societies). It appears that individualised societies were a necessary (but not sufficient) prerequisite for the evolution of culture, providing a social environment which supported the evolution of complex forms of social learning (in particular imitation). The capacity for phenotypic, cultural evolution seems correlated with particular mental capacities and social skills (see list above) which facilitated the evolution of complex forms of primate societies and primate culture. Primate social behaviour is well studied, we know less about the social life and mental capacities of non-primate species (crows, parrots, cetaceans, elephants, and others). However, when searching for animal culture, highly social animals in individualised societies are good candidates. Ants don’t imitate, they don’t learn from each other, primates do. Memes, as the replicators of culture, seem to require ‘a social host’, and memes are transmitted along social networks and depending on interactions its ‘host’ is engaged in. These seem to be the natural constraints under which culture is able to evolve in primate societies. The ‘magic numbers’ 150 and 3.8 indicate strong limitations and constraints for the future development of human societies. Systematic investigations that take these cognitive constraints into consideration could provide a basis for social agent technology that meets the cognitive demands of human primates.

Social Robots in Rehabilitation: the Case of Autism⁶

Autism

Although we use the term autism throughout this paper it is more appropriate to use the term autistic spectrum disorders (ASD) which acknowledges the fact that autism occurs in differing degrees and in a variety of forms. The National Autistic Society (NAS00) lists the following triad of impairments: 1. Social interaction (difficulty with social relationships, for example appearing aloof and indifferent to other people,

⁶This section is based on (Dau00a).

inappropriate social interactions, inability to relate to others in a meaningful way, impaired capacity to understand other's feelings or mental states). 2. Social communication (difficulty with verbal and non-verbal communication, for example not really understanding the meaning of gestures, facial expressions or tone of voice). 3. Imagination (difficulty in the development of play and imagination, for example having a limited range of imaginative activities, possibly copied and pursued rigidly and repetitively).

In addition to this triad, repetitive behaviour patterns and a resistance to change in routine can generally be observed, associated with a significantly reduced repertoire of activities and interests, stereotypical behaviour, and a tendency of fixation to stable environments. Depending on what is included in '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 can but need not be accompanied by learning disabilities. At the higher functioning end of the autistic 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' (Gra95), (GS96), (Sch97). 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. 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 feel (see books written by Temple Grandin and others), they are painfully 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 that 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 can find ways to empower them, including computer and robotic technology, so that they have the choice of whether and

to what extent they want to connect to the world of non-autistic people.

Brief project Description and Related Work

The AURORA project develops an autonomous, mobile robot as a therapeutic tool for children with autism (Dau99c), (WD99), (DW00). Conceptually, this approach is strongly related to Seymour Papert's constructionist approach towards learning (Pap80). Such an approach focuses on active exploration of the environment, namely improvisational, self-directed, 'playful' activities in appropriate learning environments ('contexts') which can be used as 'personal media'. In the mid- 1960ies Papert and his colleagues at the MIT AI LAB developed the programming language LOGO which has been widely used in teaching children. A remote controlled device (a 'turtle' robot) was developed which is moving according to a set of LOGO instructions, cf. the LEGO/LOGO Artificial Life Toolkit for children (Res89). In 1976 Sylvia Weir and Ricky Emanuel (WE76) published research which used such a LOGO learning environment to catalyse communication in an autistic child. They report on their experience with a seven-year-old autistic boy and the positive effects of his explorations in controlling a LOGO turtle on his behaviour. A more recent approach using more interactive rather than remote-controlled technology for rehabilitation of autistic children is taken in the Affective Social Quotient (ASQ) project, (Blo99). Here, embedded technology is used to support autistic children in learning about social-emotional cues. Short 'emotionally charged' video clips are used together with a set of physical stuffed 'dolls' (embodying one emotional expression) through which the child can interact with the movies. By touching the doll the child can match a doll with a video clip. A child can explore emotional situations by picking up dolls with certain emotions, or the system can prompt the child to pick up dolls that go with certain clips. A therapist is able to control and monitor the interactions. The system shows that human-intensive, repetitive aspects of existing behavioural therapy techniques can potentially be automated.

In recent years the AURORA project described below and work by Francois Michaud (Michaud, this volume; (MCLL00), (MLL⁺00)) who develops interesting interactive robotic designs, is taking up this line of work. Since end of 1998 the project AURORA (AUtonomous RObotic platform as a Remedial tool for children with Autism) investigates how an autonomous mobile robot can be developed into a remedial tool in order to encourage children to be-

come engaged in a variety of different interactions that possess features which are important elements of human social behaviour (eye-contact, joint-attention, approach, avoidance, following, imitation games etc.). The children who are interacting with the robot are between 8-12 years of age, including children who are non-verbal, i.e. they cannot use language or usually do not use language. In the rehabilitation of children with autism therapeutic issues (e.g. eye contact, joint attention, turn taking, reading mental states and emotions) are usually addressed in constrained teaching sessions (HBCH99). In contrast, robot-human interactions in the AURORA project are unconstrained and unstructured, the children are allowed to interact with the robot in whatever body position they prefer (e.g. lying on the floor, crawling, standing, cf. figure 4, they are also free to choose how they interact with the robot (touching, approaching, watching from a distance, picking it up etc.). Interference is only necessary if the child is about to damage the robot or if the child (by pressing buttons) switches off the robot so that it needs to be restarted. Such conditions are much different from other projects on robot-human interaction which are based on structured and constrained set ups (e.g. KISMET, or the ROBOTA dolls, see this volume) where the human is expected to interact with the robot while adopting a particular position and orientation towards the robot (e.g. sitting face-to-face in close distance to an interactive robot that is not moving in space). The particular challenges faced in the AURORA project, in the broader context of rehabilitation, together with a more detailed discussion of therapeutic issues involved, is given in (WD99), (DW00).

Theoretical Background and Working Hypotheses

The AURORA project deliberately uses a non-humanoid robot. based on the observation that children with autism prefer a predictable, stable environment and that many people with autism have difficulty interpreting facial expressions and other social cues in social interactions. Consequently, they often avoid social interactions since people appear unpredictable and confusing. Generally, 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 (see discussion in (Dau99c)). We hypothesise that a child with autism 1) is sufficiently interested in

‘playing’ with an interactive autonomous robot as it is used in the AURORA project, 2) the robot can engage the child in interactions which demonstrate important aspects of human-human interaction (e.g. eye-contact, turn-taking, imitation games), and 3) (as a long term therapeutic goal), while slowly increasing the robot’s behaviour repertoire and the unpredictability of its actions and reactions, the robot can be used to guide the children towards more realistic and ‘complex’ forms of social interactions resembling human-human interaction. This approach is based on two areas of theoretical work, namely mindreading and interaction dynamics. These issues and their implications for the AURORA project are described in the following two sections.

Mindreading

Generally, humans are from an early age on attracted to self-propelled objects which are moving autonomously and seemingly with ‘intention’ (Dau97). In (PP95b) a theory of human social competence is presented that consists of three units: the first unit (intentional system) identifies self-propelled movements in space and interprets them as intentional, engaged in goal-directed behaviour, such as escaping from confinement, making contact with another intentional object, overcoming gravity (e.g. seeking to climb a hill). Animate and inanimate objects are distinguished since only animate objects can move both in space and time without the influence of other objects. Movement in place is interpreted as animate but not intentional. The second unit is the social system which specifies the changes that the intentional objects undergo. It allows to interpret relations e.g. as possession or group membership. The third unit is the theory of mind system, which outputs explanation, states of mind, perception, desire, belief, and its variations. These mental states are used to explain the actions. Premack and Premack’s theory of human social competence shows great similarity with Baron-Cohen’s suggestion of four mechanisms underlying the human mindreading system (BC95). The first mechanism is the intentionality detector that interprets motion stimuli (stimuli with self-propulsion and direction) in terms of the mental states of goal and desire. These primitive mental states are basic since they allow making sense of universal movements of all animals, namely approach and avoidance, independent of the form or shape of the animal. The ID mechanism works through vision, touch and audition and interprets anything that moves with self-propelled motion or produces a non-random sound as an object with goals and desires. The second mechanism as part of Baron-Cohen’s mindreading system is the eye-direction detector (EDD) which

works only through vision. The EDD detects the presence of eye-like stimuli, detects the direction of eyes, and interprets gaze as seeing (attribution of perceptual states). This mechanism allows interpreting stimuli in terms of what an agent sees. ID and EDD represent dyadic relations (relations between two objects, agent and object or agent and self) such as 'Agent X wants Y' or 'Agent X sees Y', however they not allow to establish the link between what another agent sees and wants and what the self sees and wants. Sharing perceptions and beliefs is beyond the 'autistic universe', it requires the additional mechanisms SAM (shared-attention-mechanism, allows to build triadic representations: relations between an agent, the self, and a third object) and TOM (theory-of-mind mechanism). ID, EDD, SAM and TOM make up a fully developed human mindreading system as it exists in biologically normal children above the age of four. In normal development, from birth to about 9 months a child can only build dyadic representations based on ID and basic functions of EDD. From about 9 to 18 months SAM comes on board and allows triadic representations that make joint attention possible. SAM links EDD and ID, so that eye direction can be read in terms of basic mental states. From about 18 to 48 months TOM comes on board, triggered by SAM. The arrival of TOM is visible e.g. through pretend play. Note, that earlier mechanisms are not replaced by newer ones, they still continue to function. According to Simon-Baron's analysis children with autism possess ID and EDD. TOM is missing in all children with autism while some of them possess SAM. Referring to this theoretical framework, the working hypotheses (section 2.4) studied in the AURORA project clearly address the ID and EDD mechanisms. In the same way as biologically normal children above 4 years of age detect, are attracted to, and interpret autonomous, self-propelled objects such as robots as 'social agents', we hypothesise that children with autism can accept a mobile robot as a social agent.

Interaction Dynamics

The second strand of theories which the AURORA project is influenced by concerns interaction dynamics between babies and their caretakers as studied in developmental psychology, e.g. (Bul79), (UBKV89), (Mel96), (MM99). A more detailed account of these issues in the general context of robot-human interaction, and their relevance in the AURORA project is given in (DW00), we can only present a brief summary here. Infants seem to detect specific temporal and structural aspects of infant-caretaker interaction dynamics. It is suggested that turn-taking and imitation games allow



Figure 4: An autistic boy playing with the Labo-1 mobile robot which was kindly donated by *Applied AI Systems Inc.* The child is not afraid to let the robot come physically very close to his body, including the face.



Figure 5: The child frequently 'reaches out' to the robot, 'testing' its front sensors and eliciting the robot's response to approach or avoid. After 20 minutes the teacher ended the interaction since the child had to go back to class.

the infant 1) to identify people as opposed to other objects, and 2) to use the like-me-test in order to distinguish between different persons. Motivated by this research we suggested a conceptual framework in order to classify different and increasingly complex dynamics in robot-human interactions (DW00). Within this framework, robot-human interactions in the AURORA project are designed where synchronisation of movements, temporal coordination, and the emergence of imitation games are used as important mechanisms for making 'social contact' between the robot and the child. It is hoped that such an approach which focuses on interaction dynamics rather than cognitive reasoning mechanisms can incrementally facilitate and strengthen temporal aspects which are so fundamen-

tal to the development of social competence and the ability to socially interact with people (cf. (Hal83)).

AURORA: Preliminary Results

Initial trials in the AURORA project stressed the individual nature of the specific needs of children with autism, but they also showed that most children responded very well and with great interest to the autonomous robot, see figures 4, 5. In a recent series of comparative trials where the children were playing with the robot (condition 1) and also (separately) with a passive non-robotic toy (condition 2) children showed greater interest in interactions with the robot than with the "nanimate" toy (quantitative data will be published in a forthcoming publication by Werry and Dautenhahn). Also, children often showed increased interest in the front part of the robot where the pyro-electric sensor is attached, a sensor with strongly eye-like features (eye-like shape, located at the distal end of the robot's preferred direction of movement, prominent position raised above the chassis, direction of the sensor changing according to 'gaze'). These observations seem to confirm our hypothesis that interactions in the AURORA project can successfully built on mechanisms of intentionality detection (ID mechanism) and eye-direction- detection (EDD mechanism). Please note that mobile robots are only seen as potentially one form of therapy, which might complement other forms of therapies (see review in (DW00)). An interesting line for future research is to study the application of virtual environments for children with autism, as discussed in (Dau00a).

A particular problem we encounter in the AURORA project is that (with few exceptions) we cannot ask our subjects, they do not give verbal feedback, techniques like interviews or questionnaires are impossible. This puts particular emphasis on the analysis of behaviour and interaction. I believe that the field of socially intelligent agents has a huge potential in education, therapy and rehabilitation. However, new design and evaluation techniques and methodologies need to be developed (cf. (MJL00), (MAD+00)).

Empathy

In (Dau97) (see also (Dau99a)) I discussed empathy as a fundamental, experiential mechanism with humans use to bond and understand each other. Also, empathy can be considered as a means of social learning via bonding with other people. According to Wispe empathy is a way of 'knowing', as opposed to 'relating' which occurs in sympathy (Wis86). Brothers considers empathy (Bro89) a biological phenomenon, an 'emotional communication' system that the human *social brain*

seems to be specialised in (Bro97).

Inspired by autism research and Barrett-Lennard's cyclic/phasic model of empathy, (BL81), (BL93), I suggested in (Dau97) and (Dau99a) to distinguish between two different mechanisms: a) *empathic resonance*, an immediate, direct way of re-experiencing, and b) *biographical reconstruction*, namely reading another persons mind by re-constructing the other's autobiographical context (who that person is, where he comes from, what the relationship is with oneself, what behaviour might be expected etc.). Barrett-Lennard's empathy cycle is a process between two people, involving expressing and receiving empathy.

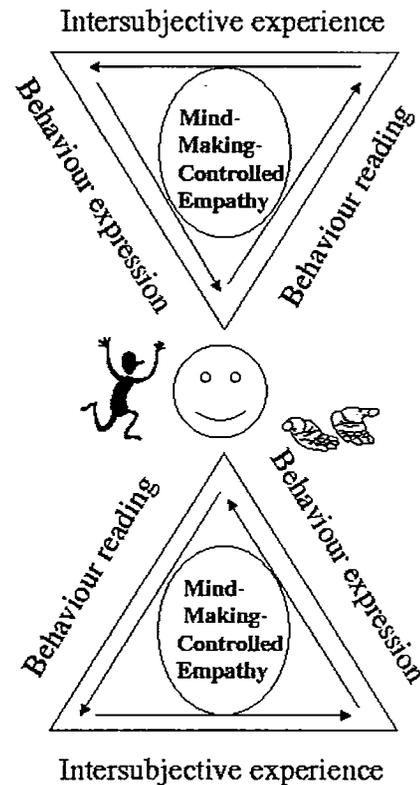


Figure 6: A Theory of empathy based on issues discussed in this paper, see text for explanation.

Recent experiments on empathic accuracy (the ability to read and understand reliably another person's intentions, beliefs, etc.), as well as neurophysiological experiments with monkeys point towards an exciting possibility how empathic resonance might actually be grounded in biological mechanisms: Neurons were found in area F5 of the monkey brain that discharge when the monkey grasps or manipulates objects, but also when the monkey observes an experimenter mak-

ing a similar gesture ((GFFR96), (RFGF98)). Arbib (Arb01) speculates that all primates (including humans) might share the mirror system as a neurobiological mechanism underlying imitation (note that imitation can be shown for humans and other apes, but is difficult to confirm for monkeys, who are nevertheless good social learners, (VF01)). Some researchers even suggest that the mirror system in F5 (analogous to Broca's area in humans, important for language), is 'grounding' language in gestures and body language (Arb01), (RA98). Although the findings of neurophysiological studies in monkeys need to be confirmed for humans and further understood for all primates and non-primate animals, it is suggested that mirror neurons could be the basis for a simulation theory of empathy (GG98). Previously, in discussions on how people ascribe mental states to themselves and others, a *simulation theory* was opposed to a *theory theory* (Gol92), (Gor92). Supporters of the simulation theory favour a process of 'putting oneself in the other's place', as opposed to (detached) reasoning about other's beliefs, emotions etc. The importance of the mirror system in this context is that (in support of the simulation theory) it could be nature's solution - at least in some primates - to solving the correspondence problem for empathy and creating *intersubjective experience* by creating a common shared context and shared *understanding* of actions and affordances. Figure 6 shows mechanisms and processes which were discussed above and how they might fit into a theory of empathy. Here, two persons are linked via a common social 'currency', namely facial expressions, body language, gestures, imitation games, interactions dynamics, spatial-temporal dynamics as they are studied in proxemics (the study of human's perception and use of space, cf. (Hal68) (Hal83)), etc. These are important for *automatic empathy* that creates *intersubjective experience* and *physiological synchrony* (LR97). This is accompanied by a cognitive, *controlled mechanism of empathy*, a more deliberative inference making (HW97), what we called biographical reconstruction. The theory sketched in figure 6 needs to be confirmed by the discovery mirror neurons e.g. for facial expressions and other gestures, as speculated in (Bro97), p. 78. Simon Baron-Cohen's theory of mindreading nicely fits in this framework, as well as Mitchell and Hamm's discussion of behaviour reading, see above. ID and EDD might (in normally developed humans) play a strong role in behaviour reading (although they do not seem to be necessary, since empathy does not rely on the visual channel alone, cf. blind people, or Mitchell and Hamm's study of behaviour reading with narratives (MH97)). TOM might be part of controlled empathy,

reasoning about another person's beliefs, desires, goals, emotions etc. and biographical reconstruction.

Empathy does not only occur in face-to-face contact with another person, it can also be evoked by reading a book or watching a movie, i.e. without any feedback from the character/person we might empathise with. If I empathise with a human being (whether real, enacted, fictional or imagined) empathy is nevertheless based on my assumption that the other human is to some extent 'like me'. An important challenge is then to create empathic relationships with non-human artifacts. Here, as shown in figure 7 we cannot assume an 'understanding' from the artifact. However, our own behaviour-reading and expression mechanisms still work, and non-biological socially intelligent agents could exploit this, see (Dau95), (BDH98), (BS00), (BA00), and other research projects on social robots described in this volume and elsewhere.

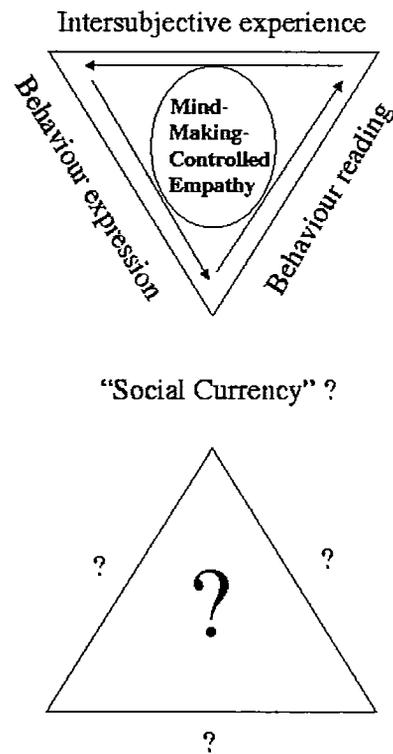


Figure 7: Empathising with agents?

Quo Vadis?

I argued in this paper that Socially Intelligent Agents (SIA) research, although strongly linked to software and robotic engineering, goes beyond a software engineering paradigm: it can potentially serve as a

paradigm for a *science of social minds*. This paper gave some indications of a few research questions that I believe are important. A systematic and experimental investigation of human social minds and the way humans perceive the social world can result in truly social artifacts, socially intelligent agents that are integrated in human society, e.g. social robots that meet the cognitive and social needs of humans. Such social agents might become more and more similar to us, in ways which could even make it difficult to distinguish between them and human beings, cf. (Fon97), (Fon00), (Dau00c).

"Once there was a robot called Jig Jag and Jig Jag lived in the countryside. One day Jig Jags lights started to flash, that meant that the robot had an idea. "I think I will go for a walk", so Jig Jag went into a field with some sheep in it and the silly robot tried to talk to the sheep, "Silly, silly, Jig Jag". Next Jig Jag saw some cows in the next field, so silly Jig Jag tried to talk to the cows! After that Jig Jag went to the shops, he wanted to buy some bolts and oil. So Jig Jag went into the hardware shop, but poor Jig Jag set the alarm off. So Jig Jag went into another hardware store across the road. So the robot tried to get into the shop but again Jig Jag set the alarm off. So poor Jig Jag had to go home empty handed." (Lauren - 8 years old, (BD99)).

Acknowledgements

I like to thank Katherine Bumby and Iain Werry whose work is summarised in this paper and helped me to develop my ideas. The AURORA project is supported by an EPSRC grant (GR/M62648).

References

Michael Arbib. The mirror system, imitation, and the evolution of language. In K. Dautenhahn and C. L. Nehaniv (Eds.), *Imitation in Animals and Artifacts*. MIT Press (in press), 2001.

Alan H. Bond and Matt Ashton. The robot as person. Workshop on Interactive Robotics and Entertainment (WIRE-2000), Pittsburgh, April 2000, 2000.

Steve Benford, Benjamin B. Bederson, Karl-Petter Akeson, Victor Bayon, Allison Druin, Par Hansson, Juan Pablo Hourcade, Rob Ingram, Helen Neale, Claire O'Malley, Kristian T. Simsarian, Danae Stanton, Yngve Sundbald, and Gustav Taxen. Designing storytelling technologies to encourage collaboration between young children. *Proc. CHI 2000*, April 1-6, 2000, The Hague, The Netherlands, 2000.

Simon Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. A Bradford Book, The MIT Press, Cambridge, London, 1995.

M. Umaschi Bers and J. Cassell. Children as designers of interactive storytellers "let me tell you a story about myself". In Kerstin Dautenhahn, editor, *Human Cognition and Social Agent Technology*, chapter 16, pages 61-83. John Benjamins Publishing Company, 2000.

Robert A. Barton and Robin I.M. Dunbar. Evolution of the social brain. In Andrew Whiten and Richard W. Byrne, editors, *Machiavellian Intelligence II Extensions and Evaluations*, chapter 9, pages 240-263. Cambridge University Press, 1997.

K. Bumby and K. Dautenhahn. Investigating children's attitudes towards robots: A case study. In *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco, available at <http://www.cogtech.org/CT99>*, pages 391-410, 1999.

A. Billard, K. Dautenhahn, and G. Hayes. Experiments on human-robot communication with robots, an imitative learning and communication doll robot. Technical Report CPM-98-38, Centre for Policy Modelling, Manchester Metropolitan University, UK, 1998.

Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, New York, Oxford, 1999.

R. W. Byrne and A. Whiten (Eds.). *Machiavellian intelligence*. Clarendon Press, 1998.

Aaron F. Bobick, Stephen S. Intille, James W. Davis, Freedom Baird, Claudio S. Pinhanez, Lee W. Campbell, Yuri A. Ivanov, Arjan Schtte, , and Andrew Wilson. The kidsroom: A perceptually-based interactive and immersive story environment. *Presence*, 8(4):369-393, August 1999.

Godfrey T. Barrett-Lennard. The empathy cycle: refinement of a nuclear concept. *Journal of Counseling Psychology*, 28(2):91-100, 1981.

Godfrey T. Barrett-Lennard. The phases and focus of empathy. *British Journal of Medical Psychology*, 66:3-14, 1993.

Susan Blackmore. *The Meme Machine*. Oxford University Press, 1999.

Katherine H. Blocher. Affective Social Quest (ASQ). Teaching emotion recognition with interactive media and wireless expressive toys. Master's Thesis for Master of Science in Media Technology Massachusetts Institute of Technology, MIT, USA, 1999.

D. Berdichevsky and E. Neunschwander. Toward and ethics of persuasive technology. *Communications of the ACM*, 42(5):51-58, 1999.

Valentin Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, 1984.

Leslie Brothers. A biological perspective on empathy. *American Journal of Psychiatry*, 146(1):10-19, 1989.

Leslie Brothers. *Friday's footprint: how society shapes the human mind*. Oxford University Press, New York, Oxford, 1997.

Jerome Bruner. The Narrative Construction of Reality. *Critical Inquiry*, 18(1):1-21, 1991.

Cynthia Breazeal and Brian Scassellati. Infant-like social interactions between a robot and a human caretaker. To appear in Special issue of *Adaptive Behavior* on Simulation Models of Social Agents, guest editor Kerstin Dautenhahn, 2000.

- M. Bullowa. *Before Speech*. Cambridge University Press, 1979.
- R. W. Byrne. Machiavellian intelligence. *Evolutionary Anthropology*, 5:172–180, 1997.
- Steven Coles and Kerstin Dautenhahn. A robotic storyteller. Proc. *SIRS2000*, 8th Symposium on Intelligent Robotic Systems, The University of Reading, England, 18-20 July 2000, pp. 393-401, 2000.
- Kerstin Dautenhahn. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16:333–356, 1995.
- Kerstin Dautenhahn. I could be you – the phenomenological dimension of social understanding. *Cybernetics and Systems*, 25(8):417–453, 1997.
- Kerstin Dautenhahn. The art of designing socially intelligent agents: science, fiction and the human in the loop. *Applied Artificial Intelligence Journal, Special Issue on Socially Intelligent Agents*, 12(7-8):573–617, 1998.
- Kerstin Dautenhahn. Embodiment and interaction in socially intelligent life-like agents. In C. L. Nehaniv, editor, *Computation for Metaphors, Analogy and Agents*, pages 102–142. Springer Lecture Notes in Artificial Intelligence, Volume 1562, 1999.
- Kerstin Dautenhahn. The lemur’s tale - story-telling in primates and other socially intelligent agents. *Proc. Narrative Intelligence, AAAI Fall Symposium 1999*, AAAI Press, Technical Report FS-99-01, pp. 59-66, 1999.
- Kerstin Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*. pages 359–374, 1999.
- Kerstin Dautenhahn. Design issues on interactive environments for children with autism. Proc. The 3rd International Conference on Disability, Virtual Reality and Associated Technologies, ICDVRAT 2000, 23-25 September 2000, Alghero, Sardinia, Italy, 2000.
- Kerstin Dautenhahn. Evolvability, culture and the primate social brain. In C. L. Nehaniv, editor, *Proceedings of the Evolvability Workshop at the Seventh International Conference on the Simulation and Synthesis of Living Systems (Artificial Life VII)*, pages 23–26. 2000.
- Kerstin Dautenhahn. Reverse engineering of societies - a biological perspective. Proc. AISB Symposium *Starting from Society - the application of social analogies to computational systems*, a symposium at AISB-00, 17th-20th April 2000, University of Birmingham, England, pp. 21-31, 2000.
- Richard Dawkins. *The Selfish Gene*. Oxford University Press, 1976.
- Kerstin Dautenhahn and Steven Coles. Robotic storytellers: A bottom-up computational framework for the study of narrative intelligence in autonomous agents. *Proc. Narrative and Interactive Learning Environments* Edinburgh, Scotland 30th August - 1st September 2000, pp. 51-59, 2000.
- D. C. Dennett. Intentional systems. *Journal of Philosophy*, 68:87–106, 1971.
- Daniel C. Dennett. *The intentional stance*. MIT Press, 1987.
- Daniel C. Dennett. The origins of selves. *Cogito*, 3, pp. 163-73, Autumn 1989. Reprinted in Daniel Kolak and R. Martin, eds., *Self and Identity: Contemporary Philosophical Issues*, Macmillan, 1991, 1989.
- Merlin Donald. Precis of origins of the modern mind: Three stages in the evolution of culture and cognition. *Behavioral and Brain Sciences*, 16:737–791, 1993.
- R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–735, 1993.
- Frans B.M. de Waal. Cultural primatology comes of age. *Nature*, 399:635–636, 1999.
- Kerstin Dautenhahn and Iain Werry. Issues of robot-human interaction dynamics in the rehabilitation of children with autism. To be published in Proc. *FROM ANIMALS TO ANIMATS*, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000), 11 - 15 September 2000, Paris, France, 2000.
- B. Edmonds and K. Dautenhahn. The contribution of society to the construction of individual intelligence. Technical Report CPM-98-42, Centre for Policy Modelling, Manchester Metropolitan University, UK, 1998.
- T. J. Eddy, G. G. Gallup, and D. J. Povinelli. Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. *Journal of Social Issue*, 49(1):87–101, 1993.
- Stan Franklin and Art Graesser. Is it an agent, or just a program?: A taxonomy for autonomous agent. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, published as Intelligent Agents III*, pages 21–35. Springer-Verlag, 1997.
- B. J. Fogg. Introduction: Persuasive technologies. *Communications of the ACM*, 42(5):27–29, 1999.
- Leonard N. Foner. Entertaining agents: a sociological case study. In W. Lewis Johnson, editor, *Proc. of the First International Conference on Autonomous Agents, Marina del Rey, CA, USA, February 5-8*, pages 122–129, 1997.
- Leonard N. Foner. Are we having fun yet? using social agents in social domains. In Kerstin Dautenhahn, editor, *Human Cognition and Social Agent Technology*. chapter 12, pages 323–348. John Benjamins Publishing Company, 2000.
- Stan Franklin. Autonomous agents as embodied ai. *Cybernetics and Systems*, 28(6), 1997.
- Jennifer W. Glos and Justine Cassell. Rosebud: A place for interaction between memory, story, and self. In Jonathon P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska, editors, *Proceedings of the Second International Conference on Cognitive Technology*, pages 88–97. IEEE Computer Society Press, 1997.
- V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119:593–609, 1996.
- V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501, 1998.

- Gerd Gigerenzer. The modularity of social intelligence. In Andrew Whiten and Richard W. Byrne, editors, *Machiavellian Intelligence II Extensions and Evaluations*, chapter 10, pages 264–288. Cambridge University Press, 1997.
- Alvin I. Goldman. In defense of the simulation theory. *Mind and Language*, 7(1-2):104–119, 1992.
- Robert M. Gordon. The simulation theory: objections and misconceptions. *Mind and Language*, 7(1-2):11–33, 1992.
- Temple Grandin. *Thinking in pictures*. Doubleday Publisher, New York, London, Toronto, Sydney, Auckland, 1995.
- T. Grandin and M. M. Scariano. *Emergence: Labeled autistic*. Warner Books, 1996.
- E. T. Hall. Proxemics. *Current Anthropology*, 9(2-3):83–95, 1968.
- E. T. Hall. *The Dance of Life: The Other Dimension of Time*. Anchor Books, Doubleday, 1983.
- Patricia Howlin, Simon Baron-Cohen, and Julie Hadwin. *Teaching Children with Autism to Mind-Read*. John Wiley and Sons, 1999.
- F. Heider and M. Simmel. An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–259, 1944.
- Michael A. Huffman. Acquisition of innovative cultural behaviors in nonhuman primates: a case study of stone handling, a socially transmitted behaviour in japanese macaques. In Cecilia M. Heyes and Bennett G. Galef Jr., editors, *Social learning in animals*, chapter 13, pages 267–289. Academic Press, 1996.
- S. D. Hodges and D. M. Wegner. Automatic and controlled empathy. In W. Ickes, editor, *Empathic accuracy*, pages 311–339. The Guildford Press, 1997.
- N. R. Jennings, K. Sycara, and M. Wooldrige. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1:7–38, 1998.
- Marc Kirschner and John Gerhart. Evolvability. *Proc. Natl. Acad. Sci.*, 95:8420–8427, 1998.
- R. W. Levenson and A. M. Ruef. Physiological aspects of emotional knowledge and rapport. In W. Ickes, editor, *Empathic accuracy*, pages 44–72. The Guildford Press, 1997.
- Jaime Montemayor, Housman Alborzi, Allison Druin, Jim Hendler, Deborah Pollack, Jessica Porteous, Lisa Sherman, Asmara Afework, Jack Best, Joe Hammer, Alexander Kriskal, Abigail Lal, Thomas Plaisant Schwenn, Lauren Sumida, and Rebecca Wagner. From pets to storykit: Creating new technology with an intergenerational design team. *Proceedings 2000 Workshop on Interactive Robotics and Entertainment (WIRE-2000) in cooperation with AAAI, April 30 - May 1, 2000, The Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania, USA*, 2000.
- Michael Mateas. Computational subjectivity in virtual world avatars. In *Socially Intelligent Agents*, pages 43–45. AAAI Press, Technical report FS-97-02, 1997.
- F. Michaud, A. Clavet, G. Lachiver, and M. Lucas. Designing toy robots to help autistic children - an open design project for electrical and computer engineering education. *Proc. American Society for Engineering Education*, June 2000, 2000.
- A. Meltzoff. The human infant as imitative generalist: a 20-year progress report on infant imitation with implications for comparative psychology. In B. G. Galef and C. M. Heyes, editors, *Social Learning in Animals: the Roots of Culture*, pages 347–370. Academic Press, New York, 1996.
- R. W. Mitchell and M. Hamm. The interpretation of animal psychology: Anthropomorphism or behavior reading? *Behaviour*, 134:173–204, 1997.
- Steven Mithen. *The prehistory of the mind*. Thames and Hudson Ltd, London, England, 1996.
- Stacy C. Marsella, W. Lewis Johnson, and C. LaBore. Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents, June 3-7, Barcelona, Spain*, pages 301–308. ACM Press, 2000.
- F. Michaud, P. Lepage, J.-D. Leroux, M. Clarke, F. Belanger, Y Brosseau, and D. Neu. Mobile robotic toys for autistic children. *Proc. International Symposium on Robotics*, Montreal, May 2000, 2000.
- A. N. Meltzoff and M. K. Moore. Persons and representation: why infant imitation is important for theories of human development. In J. Nadel and G. Butterworth, editors, *Imitation in Infancy*, pages 9–35. Cambridge University Press, 1999.
- I. Machado, C. Martinho, and A. Paiva. Once upon a time. *Proc. Narrative Intelligence, AAAI Fall Symposium 1999*, AAAI Press, Technical Report FS-99-01, pp. 115–119, 1999.
- NAS. http://www.oneworld.org/autism_uk/index.html. Last referenced on 14th of February, 2000, 2000.
- Christopher L. Nehaniv and Kerstin Dautenhahn. Living with socially intelligent agents: a cognitive technology view. In Kerstin Dautenhahn, editor, *Human Cognition and Social Agent Technology*, chapter 16, pages 415–426. John Benjamins Publishing Company, 2000.
- Seymour Papert. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, New York, 1980.
- Daniel J. Povinelli and Todd M. Preuss. Theory of mind: evolutionary history of a cognitive specialization. *Trends in Cognitive Neurosciences*, 18:418–424, 1995.
- David Premack and Ann James Premack. Origins of human social competence. In Michael S. Gazzaniga, editor, *The cognitive neurosciences*, pages 205–218. A Bradford Book, The MIT Press, 1995.
- D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4:515–526, 1978.
- T. Quick, K. Dautenhahn, C. Nehaniv, and G. Roberts. The essence of embodiment: A framework for understanding and exploiting structural coupling between system and environment. *Proc. CASYS'99, Third International Conference on Computing Anticipatory Systems*, HEC, Lige, Belgium, August 9 -14, 1999.

- T. Quick, K. Dautenhahn, C. Nehaniv, and G. Roberts. On bots and bacteria: Ontology independent embodiment. Proc. ECAL99, Fifth European Conference on Artificial Life, Switzerland, September, pp. 339-343, 1999.
- T. Quick, K. Dautenhahn, C. Nehaniv, and G. Roberts. Understanding embodiment, system-environment coupling and the emergence of adaptive behaviour. In: A. Drogoul, J.-A. Meyer (eds.) *Intelligence artificielle située*, Hermes Science Publications, Paris, 1999, pp. 13-31, 1999.
- G. Rizzolatti and M. A. Arbib. Language within our grasp. *Trends in Neurosciences*, 21(5):188-194, 1998.
- Mitchel Resnick. LEGO, LOGO, and Life. In C. G. Langton, editor, *Proc. of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, Los Alamos, New Mexico, September 1987*, pages 397-406, 1989.
- G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognition and Brain Research*, 3:131-141, 1998.
- Simon M. Reader and Kevin N. Laland. Do animals have memes? *Journal of Memetics - Evolutionary Models of Information Transmission*, 3(2), 1999.
- Stephen John Read and Lunn Carol Miller. Stories are fundamental to meaning and memory: for social creatures, could it be otherwise? In Robert S. Wyer, editor, *Knowledge and Memory: the Real Story*, chapter 7, pages 139-152. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995.
- B. Reeves and C. Nass. *The Media Equation*. Cambridge University Press, 1996.
- Robert Jay Russell. *The Lemurs' Legacy. The Evolution of Power, Sex, and Love*. G.P. Putnam's Sons, New York, 1993.
- Susanne Schäfer, editor. *Sterne, Äpfel und rundes Glas. Mein Leben mit Autismus*. Verlag Freies Geistesleben und Urachhaus GmbH, Stuttgart, Germany, 1997.
- Phoebe Sengers. Do the thing right: an architecture for action-expression. In Katia P. Sycara and Michael Wooldridge, editors, *Proc. of the Second International Conference on Autonomous Agents, Minneapolis/St. Paul, USA, May 9-13*, pages 24-31, 1998.
- Phoebe Sengers. Narrative intelligence. In: *Human Cognition and Social Agent Technology*, Ed. Kerstin Dautenhahn, John Benjamins Publishing Company, 2000.
- Paul W. Sherman, Jennifer U.M. Jarvis, and Richard D. Alexander, editors. *The Biology of the Naked Mole-Rat*. Princeton University Press, Princeton, N.J., 1991.
- S. Tseng and B. J. Fogg. Credibility and computing technology. *Communications of the ACM*, 42(5):39-44, 1999.
- M. Tomasello, A.C. Kruger, and H.H. Ratner. Cultural learning. *Behavioral and Brain Sciences*, 16(3):495-552, 1993.
- I. Uzgiris, J. Benson, J. Kruper, and M. Vasek. Establishing action-environment correspondences: Contextual influences on imitative interactions between mothers and infants. In J. Lockman and N. Hazen, editors, *Action in Social Context*, pages 103-127. New York: Plenum, 1989.
- Elisabetta Visalberghi and Dorothy Frigaszy. Do monkeys ape? - ten years after. In K. Dautenhahn and C. L. Nehaniv (Eds.), *Imitation in Animals and Artifacts*. MIT Press (in press), 2001.
- Stuart N. K. Watt. Seeing things as people: anthropomorphism and common-sense psychology. Unpublished PhD thesis, Department of Psychology, The Open University, 1997.
- Iain Werry and Kerstin Dautenhahn. Applying robot technology to the rehabilitation of autistic children. Proc. SIRS99, 7th International Symposium on Intelligent Robotic Systems '99, 1999.
- S. Weir and R. Emanuel. Using LOGO to catalyse communication in an autistic child. Technical report, DAI Research Report No. 15, University of Edinburgh, 1976.
- Andrew Whiten and R. W. Byrne (Eds.). *Machiavellian intelligence ii.: Evaluations and extensions*. Cambridge University Press, 1997.
- A. Whiten, J. Goodall, W.C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C.E.G. Tutin, R.W. Wrangham, and C. Boesch. Cultures in chimpanzees. *Nature*, 399:682-685, 1999.
- Andrew Whiten. *Natural Theories of Mind*. Basil Blackwell, 1991.
- Lauren Wispe. The distinction between sympathy and empathy: to call forth a concept, a word is needed. *Journal of Personality and Social Psychology*, 50(2):314-321, 1986.
- M. Wooldridge, N. R. Jennings, and D. Kinny. The gaia methodology for agent-oriented analysis and design. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(3):285-312, 2000.