

Ants don't have Friends – Thoughts on Socially Intelligent Agents

Kerstin Dautenhahn

Department of Cybernetics, University of Reading
Whiteknights, PO Box 225
Reading RG6 6AY, United Kingdom
kd@cyber.reading.ac.uk

Introduction

The question ‘what is an agent?’ has been under discussion for many years. However, a consensus exists that the term ‘agent’ only makes sense in a *multi-agent* context, namely if there are at least two agents, and assuming interaction and/or communication between the agents. Agent research is generally done fairly independently in different research areas, separated by the nature of the agents (natural or artificial). This paper presents some thoughts on agency and sociality. Social intelligence is studied in the context of human-style forms of social behaviour. Issues like embodiment, believability, rationality, social understanding, and different levels of social organisation and control are discussed.

Agents

Since the term ‘agent’ is very diversely used in literature, does it make sense to discuss a terminology with respect to social agents? Throughout this paper I try to concretise terms of ‘agency’ and ‘social intelligence’, taking a particular ‘human’ stance, i.e. focusing on human forms of social intelligence.

Let us for the purpose of this paper define agents as entities inhabiting our world, being able to react and interact with the **environment** they are located in and other agents of the same and different kind. Thus, agents are situated and embedded in a ‘habitat’, they act by using resources from this environment and therefore change the environment. Full or partial *autonomy* and control about issues which are crucial for the existence of an agent (e.g. energy, space), i.e. maintaining and controlling the relationships to the environment, are considered to be important. A book standing on book shelves is not an agent. It’s form is adapted to its environment, it uses resources (space on the shelf), reacts e.g. to temperature and light (burns when having contact to heat or it’s pages turn yellow when being exposed to sunlight, and objects of similar kind (other books) are close to it. The book also dynamically changes state over time, without explicit external forces (the pages age, fall apart after a long period of time). However, the book itself is not actively

influencing it’s state, it is not modifying the relationship to its environment. We never consider a book to be an agent. A dog is an agent, a crow, a lemur, a child. Why? This paper tries to approach the answer.

A brief history of agency

What kind of agents do inhabit our world? For millions of years **biological agents** does exist, plants and animals. These agents consist of single cells, form aggregations and colonies, form complex entities by single cells merging together or divide and specialise to form multi-cellular organisms. For thousands of years specific animal agents, humans, have been interested to build artefacts which are imitating or depicting biological agents, paintings and puppets, made of stone, clay, paper, or synthetic media, paintings or statues depicting prey, livestock, other humans. These artefacts have been used as religious objects, luxurious gifts, efficient tools, ordinary toys. It happened (for reasons still under discussion) that humans are above all social animals (Aronson 1994), they survive in groups, form societies and culture, learn by tradition and education, divide labour, trade, and enjoy the company of other humans being close to them. The social intelligence hypothesis states that human intelligence “originally evolved to solve social problems and was only later extended to problems outside the social domain” ((Byrne 1995), (Byrne & Whiten 1988), (Dautenhahn 1995)). Thus, mental occupation with social dynamics could have paved the way towards abstract thinking and logic. Even if our ancestors might have lived in a fairly predictable environment (e.g. as gorillas did before human intervention) group members are never totally predictable, they require constant reconsideration, re-assessing and re-learning of relationships and group structures. Humans share sociality with other biological agents, like social insects (e.g. bees and termites), species of birds like parrots and crows, whales and dolphins. Humans possess an elaborated and efficient communication system, language, which might have developed as an efficient means for social grooming, it allows to communicate about issues on different level of abstractions but is less immediate

than communication by ‘body language’ and facial expressions. Humans still spend 60% of their time on discussing about other people and social matters (Dunbar 1993).

As biological agents humans are specifically attracted to ‘life’, watching and studying and talking to other biological agents. Humans seem to be naturally biased to perceive self-propelled movement and intentional behaviour ((Premack & Premack 1995), (Dennett 1987)), to perceive and recognise other biological agents. Humans are not the only tool-designers in the animals world, but they happen to be the best ones, in terms of creativity of using material and functionality of the results. For a few year humans have been developing specific agents based on silicon. Part of these **artificial agents** are made of software, *computational agents* which can take different forms, called ‘mobile agents’ when navigating networks, called ‘intelligent agents’ when they solve tasks which humans did before, they assists humans e.g. by dealing with boring and/or repetitive tasks like searching the Web or databases, filtering email etc. Many of these computational agents do not become visible to the human user as independent entities, e.g. they act in the background and their existence is, once activated, only visible in terms of effects and functionalities.

It turned out that in cases when computational agents considerably interact with humans (e.g. in entertainment applications, or as personal agents) that humans want them to appear believable. The idea is attractive that humans should interact with agents in a *natural* way. *Believable agents* (Bates 1994) give an ‘illusion of life’. They need not necessarily appear or act like biological agents, but some aspect in their behaviour has to be natural, appealing, life-like. Research in believable agents benefits significantly from animation work and artistic skills to creature fictional, imaginary but believable creatures.

A parallel development has occurred in the development of physical tools, *robots*. For decades ‘life-like’ robots have only been known from science fiction literature. Robotic systems existed in production lines or other areas outside human society. However, currently robots are already acting autonomously in human-inhabited environments (service robots, e.g. as floor-cleaning devices or assistants in hospitals), ongoing research aims at enhancing autonomy and improving the robot-human interface, making robots ‘friendly’, believable. Cooperative and collective behaviour has been studied with these physical artificial agents, namely robotic agents. Since humans are so naturally bonded to biological agents, their artificial counterparts, too, are to become part of human life, part of human culture. Such creatures might be considered as a new species, artificial agents which are treated similarly to biological agents and might partly take their roles.

Thus, research on computational and robotic agents

have steadily converged towards common issues in a domain where an important part of the functionality of the artificial agents is interaction with humans. Issues of agency, believability and sociality are examples for common research issues. These issues go back to mechanisms which are basically natural, like perception, communication. Thus, learning about artificial is coupled to learning about life. On the other hand, the study of biological life and living can further research on artificial agents. It is in this particular context characterised by an overlapping of the domains of biological, computational and robotic agents that the question arises whether a common ‘social interface’ might be thought of, either as a conceptual construct or a technical implementation.

Embodiment. Embodiment is naturally given in biological agents, but under discussion in the domain of artificial agents. All biological system have a body, they are living through their body, their existence cannot be separated from it. The issue of embodiment has recently attracted particular attention, opposed to ‘classical AI’ (mainly confined to human problem solving which is modelled as the internal manipulation of symbols representing items in the real world) the new direction is called ‘Embodied AI’. EAI stressed the need to study intelligence in an embodied system. The emphasis on physical agents led to cognitive robotics (Brooks 1996). Recently the question came up what embodiment can mean to a software agent (Kushmerick 1997).

Evidence from studies of humans give overwhelming arguments that humans can only act and react through their body, that it does not make sense to conceive the human ‘mind’ being separable from the body. Thus, the issue *that* embodiment matters for intelligence, life and agency is nowadays widely accepted. But the question how and to what extent embodiment matters is still open. Is a software environment in which computational agents ‘lives’ comparable to the environment biological agents are living in? Can we compare complex ecosystems like the tropical rainforest or the Namib Desert which biologists still seek to understand in all its complex and interconnected dimensions, with the memory space inside a computer? Can inputs (e.g. keyboard commands) and actions (e.g. unix commands) be compared to the sensori-motor system of animals? Have flocks of birds migrating from Scandinavia to Africa anything in common with mobile software agents navigating the internet? The scientific discussion is still open. The danger to end up in frameworks based on pure metaphorical comparisons is obvious.

In the rest of this paper I discuss a few concepts about the specific, complex way *humans* understand the world. If we want artificial agents as social interaction partners for humans then they have to be a bit ‘like us’.

Human Agents

Remembering and Understanding in Stories. Roger C. Schank and Robert P. Abelson point towards the relation of stories to knowledge and memory and the role of stories in individual and social understanding processes (Wyer 1995). They introduce the hypothesis that “stories about one’s experiences and the experiences of others are the fundamental constituents of human memory, knowledge, and social communication”. They emphasise that new experiences are interpreted in terms of old stories. Remembering static ‘facts’ about objects or ourselves (telephone numbers, addresses, names etc.) are the results, but not the basic units of remembering processes. Remembering can in this way be thought of as a process of creating and inter-relating stories, constructing and re-interpreting new stories on the basis of old ones, using our embodied ‘self’ as the point of reference.

Evidence about the structure of human memory, namely that mechanisms of remembering, perceiving and re-interpreting the world – in particular the social world – is mainly based on ‘stories’, might give us an explanation for the daily-life experience that humans seem to be addicted to stories! Humans enjoy throughout their whole life reading, watching, telling, inventing and enacting stories. They read novels, fairytales, science-fiction literature, they watch movies on TV, in cinema, they enjoy theatre plays, etc. Humans spend most of their spare time enjoying stories. Technology (e.g. books, video tapes, CD-ROMs) gives us more and more efficient means of preserving, reusing, inventing stories about history, science, culture itself, both on the level of societies as well as on the level of individual persons.

The Autobiography. The behaviour and appearance of any biological agent can only be understood with reference to its *history*. The history comprises the evolutionary aspect (phylogeny) as well as the developmental aspect (ontogeny). Social behaviour can then only be understood when interpreted in its *context*, considering past, present and future situations. This is particularly important for life-long learning human agents who are continuously learning about themselves and their environment and are able to modify and their goals and motivations. Using the notion of ‘story’ we might say that human are constantly telling and retelling stories about themselves and others. Humans are *autobiographic agents*. The way how humans remember and understand the world seems to be consist of constructive remembering and re-collection processes in a *life-long perspective*, i.e. referring to the autobiographic aspect as an ongoing re-construction of the own history and creating the concept of individual personality. The approach to focus on the life-long aspect of human memory is in line with research which has been carried out in psychology on *autobiographic memories*. In order to account for this autobiograph-

ic aspect of the individual I suggest the concept of an **autobiographic agent** as an embodied agent which dynamically reconstructs its individual ‘history’ (autobiography) during its life-time.

Rationality. Rationality plays a particular role in the daily life of most of us. We are encouraged to consider problems in a ‘rational’ manner as opposed to react spontaneously and ‘emotional’. Tools and technology based on logical, rational principles further strengthens the bias to think and behave rationally. Social conventions and rules are regulating our life in order to maintain the social system. Thus, is maximizing rationality what we should go for? Antonio Damasio’s case studies on humans who (after specific damages of brain areas) completely preserved their rationality and general ‘intelligence’ (Damasio 1994) suggest that this might not be desirable. His patients are no longer social beings, neither are their decisions ‘successful’ in terms of furthering their conditions of living. They cannot survive properly, they are no longer adapted to their environment. They completely preserved their ‘knowledge’ about the world, i.e. they still know the rules and the set of behaviours which are appropriate in certain situations. Their ‘knowledge’ and analytic skills are unimpaired, but they seem to lack the ability to apply them properly when they are confronted to real world situations. These humans seem to have lost their skill of becoming *engaged*, being embedded in a situation, a sense of *belonging* to the world. This implies that human decision-making cannot sufficiently be defined by knowledge and applying rules, humans rather ‘feel’ the right decision, i.e. judging alternatives by means of experiential understanding, reconstructing and relating alternatives to their own bodily ‘existence’ as the point of reference. Damasio’s patients show a typical indifference towards their decisions. But decision making only matters when the potential outcomes have a *meaning* to the decision maker. This does not imply that human decision-making and problem-solving is completely determined by internal, experiential processes, emotions etc. It only suggests that these internal dynamics provide the link to the real world, they form the frame of reference in which alternatives are being judged according to their meaning to a human being. To put it shortly: without experiential understanding the world is meaningless, it does not matter for a rational agent whether to take one or the other alternative in a given situation. For an embodied agent it matters whether to survive or not, whether to live or die, whether to be happy or unhappy. Intelligence is related to a living organism, the body is not a container for cognition but a means for being able to relate to the world.

Thus, going back to our original question about the role of rationality in our life: rationality is useful, but it is most successful when being grounded in a living body.

Social Understanding. What is happening when humans understand each other? Is it more a rational process of matching and manipulating symbols and representations about beliefs, desires and intentions which can be decoupled from the phenomenological dimension of being in the world? Given the notions of ‘stories’ and ‘autobiographic agents’, can social understanding be modelled by the process of matching representation of scripts and stories about others? Or is (social) understanding grounded in an embodied re-experiencing? And if it is, what are the relevant processes involved in the latter? What is the role of stories and the autobiography in this process?

In (Dautenhahn 1997) I discuss the phenomenological dimension of social understanding, I can only give a short summary here. Basically, I suggest that social understanding is based on empathy and in particular emphatic resonance as an experiential, bodily phenomenon of internal dynamics, and on a second process, the biographic re-construction which enables the empathising agent to relate a concrete communication situation to a complex biographical ‘story’ which helps to interpret and understand social interactions. I consider the internal dynamics of empathic resonance a basic mechanism of bodily, experiential grounding of communication and understanding. A state of willingness and openness towards another embodied, dynamic system is a direct, immediate way of relating to another person and becoming engaged in a communication situation. This might be a necessary condition for synchronised coordination processes (e.g. in verbal and non-verbal communication), and a prerequisite of ‘true’ social understanding. In my view, social understanding between humans requires the coupling of the behavioural level (external dynamics, interaction and communication) with the internal dynamics of empathic, experiential understanding.

Social organisation and control. Natural evolution of biological social agents demonstrates two impressive alternatives of sociality, namely *anonymous* and *individualised* societies. Social insects are the most prominent example of anonymous societies. Group members do not recognise each other individually. If we remove a single bee from a hive no search behaviour is induced: ants don’t have friends¹. The situation is quite different in individualised societies where primate societies belong to. Here the individual recognition gives rise to complex kinds of social interaction and the development of various forms of social relationships. On the behavioural level social bonding, attachment, alliances, dynamic (not genetically determined) hierarchies, social learning etc. are visible signs for individualised societies. The evolution of language, spreading of traditions and the evolution of culture are further developments of individualised societies.

¹Thanks to Rodney Brooks for this phrase.

Within a society a human being is part of social on different levels of social organisation and control. An individual is itself integrated insofar as if it consists of numerous components, subsystems (cells, organs) whose survival is dependent on the survival of the system at next the higher level. In the case of eusocial agents (e.g. social insects and naked mole rats) a genetically determined control structure of a ‘superorganism’ has emerged, a socially well integrated system. The individual itself plays no crucial role, social interactions are anonymous.

Many mammal species with long-lasting social relationships show an alternative path towards socially integrated systems. Primary groups, which typically consist of family members and close friends, emerged with close and often long-lasting individual contacts. Primary groups are here considered a network of ‘con-specifics’ who the individual agent uses as a testbed and as a point of reference for his social behaviour. Members of this group need not necessarily be genetically related to the agent. Social bonding is guaranteed by complex mechanisms of individual recognition, emotional and sexual bonding. This level is the substrate for the development of social intelligence where individuals build up shared social interaction structures, which serve as control structures of the system at this level. Even if these bonding mechanisms are based on genetic predispositions, social relationships develop over time and are not static. I proposed in a previous section to use the term ‘autobiographic agent’ to account for the dynamical aspect of re-interpreting the agent’s (social) ‘history’.

Larger, higher level groups emerge by additional control structures. Humans might have an upper limit of 150 for the size of groups with mainly face-to-face interaction and communication. According to (Dunbar 1993) 150 might be the cognitive limit with whom one person can maintain stable relationships, as a function of brain size. Larger groups of people can be handled by control mechanisms like adopting roles which can be indicated by symbolic markers (uniforms, batches), or stereotypical ways of interaction (e.g. rules for greeting each other, or templates for writing and answering letters). Higher level control structures are not simply enslaving or subsuming the lower levels in the way how the organism as a system is ‘enslaving’ its components (organs, body parts). The individual which is as a social being embedded in primary groups, does not depend completely on the survival of a specific system at a higher level, the dependency is more indirect than in the case of social insects or the organ-body relationships. Secondary and tertiary levels have mutual exchanges with the level of the individual, social, autobiographic agent.

The Social Interface Is it possible and desirable to construct a common ‘social interface’ as a technical or conceptual construct in which different ‘species’

of agents can become engaged? Software agents and physical agents (robots) need not necessarily have a 'natural' form of social behaviour, communication and interaction, they can build up social structures within their own communities. Aspects of believability or experiential social understanding need not necessarily play a role in software or robotic agent societies. A variety of social structures might emerge (hierarchies, formation of subset, 'dialects' of communication and interaction within larger groups etc.), influenced by domain specific requirements and constraints. Specific dynamics and phenotypes of interaction can result from the selected communication channels, the chosen protocols, and the specific processing and implementation details. But interactions (e.g. in communication situations or cooperative task solving) with humans create a need for all these creatures to behave 'naturally', i.e. in a way which is acceptable and comfortable to humans, so that the human user or collaborator can accept artificial agents as companions or 'interaction partners'. The *social interface* is therefore a specific 'context', a *physical or virtual human-inhabited space* where verbal or non-verbal cross-species interactions occur. Creating a cross-species society of 'socially intelligent agents' could provide a testbed to study issues like embodiment, as discussed above.

A vision of society?

Humans adapt to technology, human cognition is shaped by behaviour, appearance, means of interaction and communication with artificial agents in frequent daily encounters. Every act of social encounter has an element of mutual adaptation, however it could be that one partner is constantly more adapting than the other. Humans are experts in learning and adaptation, they can very flexibly get used to even the most awkward interface (e.g. command-line control of a computer or programming a VCR). But it seems to be desirable that artificial agents primarily adapt to human needs and human ways of interaction and living, and not vice versa. Designing artificial agents which make interaction natural for humans is necessary if humans should not act (and think) like artificial agents. But humanising the interface depends on what is considered as being 'human'.

Violent Apes. I now like to discuss one aspect of human-style social intelligence which is as real as important to our life and survival, namely violence. Humans are above all social animals, and they are violent ones. Richard Wrangham and Dale Peterson discuss in 'Demonic males' (Wrangham & Peterson 1996) violence in the context of human evolution. They suggest that the violent 'temperament' originates in a specific form of social organisation which the ancestors of the human and the chimpanzee species had in common and which has persisted until today. No matter of whether one agrees to the argumentation given in the book, it nevertheless points toward the aspect of

violence which is deeply part of human society. A variety of partly highly complex control strategies and mechanisms have evolved in different human cultures, but physical violence and in particular warfare is still part of our life, and a prominent part in many countries in the world. Additionally, non-physical violence is even more widespread, and here, too, different psychological or behavioural strategies have been developed to control it. If we call the situation described so far 'realistic', is then research on socially intelligent agents which are intended to be the user's friend, to help and assist, to make his/her life easy, and to further social contact with other people (e.g. Web agents finding 'equally-minded' people), are these more positively, 'peaceful' oriented visions of human sociality appropriate or rather naive? The assumptions are not that unrealistic, the fact that humans have different interests and goals, do not want to give access to their knowledge and personal information to the general public, and have to 'trust' their interaction partners is prominent part of agent research. Thus, agent research is (based on models from social sciences, sociology etc.) using fairly realistic assumptions about human social behaviour. On the other hand: Humans can to a great extent chose how they want to lead their lives, they are also able to change their whole social system within a few years. The fact that at present violence still plays an important part in our society does not necessarily mean that the same is true in 200 years time. It can become worse, or better, or stay at it is. But societies can change, and technology has always played an important part in these transitions, in particular technologies to control people (weapons) and means of communication (like telephone, email).

Conclusion

The following terminology summarises the points which I addressed in this paper and which might be useful as a basis for discussions.

- **Embodiment.** Embodiment means the structural and dynamic coupling of an agent with its environment, comprising external dynamics (the physical body embedded in the world) as well as the phenomenological dimension, internal dynamics of experiencing and re-experiencing. Both kinds of dynamics are two aspects emerging from the same state of being-in-the-world.
- **Agents.** Agents are entities inhabiting our world, being able to react and interact with the environment they are located in and other agents of the same and different kind.
- **Biological agents.** Animals and plants, both their behaviour and appearance can only be understood with reference to the historical context, their phylogeny and ontogeny. In biological agents issues of aliveness, autonomy and embodiment are inseparably interconnected in a complete system.

- **Artificial agents:** Artificial agents are made by man rather than nature. Current technology is silicon-based and we can distinguish computational and physical, robotic agents.
- **Believable agents.** Believable agents are artificial agents which are built for being presented to humans as ‘characters’ (opposed to intelligent agents which can act in the background). They appear ‘life-like’, humans find them appealing and interesting and can develop a personal relationship to them. Biological agents are genuinely believable, since they *are* alive instead of simulating life.
- **Autobiographic agents.** They can be defined as embodied agents which dynamically reconstructs their individual autobiographical ‘story’ during their lifetime. This autobiography reflects stories about the agents themselves as well as encounters and relationships with other agents.
- **Human social intelligence.** Humans live in individualised societies, individuals interact as ‘persons’, their coupling with the world consists of external (behavioural, structural) aspects as well as experiential, empathic aspects of internal dynamics. Human-style social intelligence can be defined as an agent’s capability to develop and manage relationships between individualised, autobiographic agents which, by means of communication, build up shared social interaction structures which help to integrate and manage the individual’s interests in relationship to the interests of the social system at the next higher level. The term *artificial social intelligence* is then an instantiation of human-style social intelligence in artificial agents. Thus, I use the term social intelligence always in the context of human-style social interaction and behaviour. A single individual belonging to a social insect colonies would therefore not be considered as a socially intelligent agent, because its intelligence is routed in the colony, the superorganism.
- **Socially intelligent agents (SIA).** Socially intelligent agents are biological or artificial agents which show elements of (human-style) social intelligence. This social intelligence can be natural (humans) or artificial (computational agents and robotic agents).

Exploitation? What happens if socially intelligent agents become so ‘familiar’ and ‘natural’ that humans develop a deep personal relationship to them? In the area of software pets discussions have already emerged whether these artificial life forms exploit human instincts for caring and nurturing (the creatures can ‘die’ if the human does not treat them properly). Just as children can become ‘addicted’ to these software pets, adults could become addicted to certain technologies. People might resist to switch off a computer or quit a program, causing the same ethical problems as in animal welfare concerns. But is this anything new? I

believe not. Humans are life-long learners, they are curiously investigating the world, searching for adventures, entertainment, intellectual challenges, in particular if they are presented in the form of ‘good stories’ (see above). Humans develop a variety of preferences, and some people like technology and (have already today) developed a special relationship to such products. In the same way people have developed for thousands of years special relationships to a tool, a piece of jewellery, a puppet, a painting or a car. Developing individual preferences and becoming ‘engaged’ is human, it is basically natural.

References

- Aronson, E. 1994. *The social animal*. New York: W.H. Freeman and Company.
- Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* 37(7):122–125.
- Brooks, R. 1996. Behavior-based humanoid robotics. In *Proc. 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 96*, 1–8.
- Byrne, R. W., and Whiten, A. 1988. *Machiavellian intelligence*. Clarendon Press.
- Byrne, R. 1995. *The thinking ape, evolutionary origins of intelligence*. Oxford University Press.
- Damasio, A. R. 1994. *Descartes’ Error. Emotion, Reason and the human brain*. New York: G.P. Putnam’s Sons.
- Dautenhahn, K. 1995. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems* 16:333–356.
- Dautenhahn, K. 1997. I could be you – the phenomenological dimension of social understanding. *Cybernetics and Systems* 25(8):417–453.
- Dennett, D. C. 1987. *The intentional stance*. MIT Press.
- Dunbar, R. I. M. 1993. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences* 16:681–735.
- Kushmerick, N. 1997. Software agents and their bodies. *Minds and Machines* 7(2):227–247.
- Premack, D., and Premack, A. J. 1995. Origins of human social competence. In Gazzaniga, M. S., ed., *The cognitive neurosciences*. A Bradford Book, The MIT Press. 205–218.
- Wrangham, R., and Peterson, D. 1996. *Demonic males: apes and the origins of human violence*. Houghton.
- Wyer, R. S. 1995. *Knowledge and memory: the real story*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.