

The Art of Designing Socially Intelligent Agents – Science, Fiction and the Human in the Loop

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

Abstract

In this paper *socially intelligent agents* (SIA) are understood as agents which do not only from an observer point of view behave socially but which are able to recognize and identify other agents and establish and maintain relationships to other agents. The process of building socially intelligent agents is influenced by what the human as the designer considers ‘social’, and conversely agent tools which are behaving socially can influence human conceptions of sociality. A Cognitive Technology (CT) approach towards designing SIA affords as an opportunity to study the process of 1) how new forms of interactions and functionalities and use of technology can emerge at the human-tool interface, 2) how social agents can constrain their cognitive and social potential, and 3) how social agent technology and human (social) cognition can co-evolve and co-adapt and result in new forms of sociality. Agent-human interaction requires a cognitive fit between SIA technology and the human-in-the-loop as designer of, user of, and participant in social interactions. Aspects of human social psychology, e.g. storytelling, empathy, embodiment, historical and ecological grounding can contribute to a believable and cognitively well-balanced design of SIA technology, in order to further the relationship between humans and agent tools. It is hoped that approaches to believability based on these concepts can avoid the ‘shallowness’ that merely take advantage of the anthromorphizing tendency in humans. This approach is put into the general framework of Embodied Artificial Life (EAL) research. The paper concludes with a terminology and list of guidelines useful for SIA design.

1 Introduction

What are *socially intelligent agents* (SIA)? How are they related to biological agents, intelligent agents, software agents, or robotic agents? What are *agents* to begin with? The term agent is often used to refer to very different entities, like computational units, software tools or ‘life-like’, believable agents. Three different approaches to defining agents that have been proposed include a common sense, a formal, and a computational account.

- Dictionaries (e.g. [HC80]) generally list two different meanings of *agent*: 1) a person who acts for, or who manages the business affairs of, another or others, 2) a person used to achieve something, to get a result. Both meanings presuppose that agents a) have a purpose, a goal, b) that other agents exist, c) that there is an underlying relationship between at least two of these agents. And, last but not least, agents are persons. Thus, the common sense meaning of agent is always related to persons who solve a task as a representative of or authorized by another person.
- The term ‘agent’ is often used in a multi-agent context, however, Luck and d’Inverno present an interesting account which applies to single objects: in [Ld95] they formally define *objects* with a set of actions and attributes, *agents* (objects with goals) and *autonomous agents* (agents with motivations which produce goals). In their definition agency is transient, e.g. a cup can have a goal (a container for coffee), but this goal is attributed by a human in a specific context and not intrinsic to the object. Thus, agency is not internal but attributed, it is in the mind of the observer. An object is interpreted as a agent, it cannot ‘be’ an agent. We can put it differently and say that the environment can *afford* agents. The case of autonomous agents is then different, since in then agency is encapsulated within the agent and need not be attributed from the outside, as goals are produced by an underlying motivational system.
- In [FG97] Stan Franklin and Art Graesser discuss different definitions of agents in agent research and formally define autonomous agents as follows: “An *autonomous agent* is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future”. This computational definition, and their taxonomy of autonomous agents which comprises both artificial and biological agents, attempts to discuss agency and autonomy in a cross-disciplinary sense, namely conceiving computational, robotic and biological agents as three different kingdoms within the autonomous agents’ taxonomy.

In the context of socially intelligent agents, we elaborate on such definitional considerations by addressing essential features of social intelligence of agents in the natural and artificial world. Below, the paper outlines characteristics of human social intelligence, its implications for SIA design, and the interrelationship between human social intelligence and SIA. Moreover, the concepts which are discussed in this paper characterize the psychological basis of believability. Story-telling, empathy, historical grounding (autobiography), and ‘ecological grounding’ in an environment are identified as factors which are relevant to the way how humans understand the (social) world. This argument is illustrated with examples, e.g. Cyberpet technology. It is hoped that approaches to achieving believability based on these concepts can avoid the ‘shallowness’ and ‘cheating’ of approaches to believability that merely take advantage of the anthromorphizing tendency in humans.

SIA can be discussed in the general context of technology which aims at the design of systems which are ‘intelligent’ (or behave intelligently) or show life-like behavior. The former is the major research issue in artificial intelligence (AI) research, while the latter is addressed in artificial life (AL) research. I argue that conceptions of intelligence and life can hardly be defined objectively, and consider ‘intelligence’ and ‘life’ as concepts which are located not inside a specific natural or artificial system but which are rather constructed and attributed by humans a) in a certain context by the process of interaction and understanding, or b) between humans in processes of agreement and the formation of conventions in the social environment in which these humans are embedded in. I discuss the potential contribution of Artificial Life (AL), which studies the emergence of complexity resulting from interactions between systems, to SIA research. Two AL research directions are outlined: a) the quest for a logic of life, opposed to b) studying the natural form of complexity in artificial media by constructing systems. The latter is the general research agenda of *Embodied AL* ([Mae90], [SB95]). Its starting points for constructing systems are the physical properties of the matter, not abstract formalisms. I interpret concepts like ‘believability’, ‘stories’ and ‘social understanding’ within this framework of creating ‘life-like’ artifacts. EAL might thus provide a valuable framework for the study of different cognitive aspects and implications of SIA technology. This approach is linked to the field of Cognitive Technology (CT), see [GMM97]. Cognitive Technology is defined as follows:

Cognitive Technology (CT) is the study of the integrative processes which condition interactions between people and the objects they manipulate. It is concerned with how technologically constructed tools (A) bear on dynamic changes in *human perception*, (B) affect natural *human communication*, and (C) act to control human cognitive adaptation. Cognitive systems must be understood not only in terms of their goals and computational constraints, but also in terms of the external physical and *social environments* that shape and afford cognition. Such an understanding can yield not only technological solutions to real-world problems but also, and mainly, tools designed to be sensitive to the cognitive capabilities and *affective characteristics* of their users. ([MNG97]).

According to this definition CT has to understand human perception, communication, social and affective constraints in order to optimize cognitive fit of technologically constructed tools. A CT approach towards designing SIA affords to study the process of how new forms of interactions, functionalities and use of technology can emerge at the human-tool interface. These kinds of interaction need not necessarily mimic nature and copy ‘natural’ forms of interaction. Instead, there can emerge qualitatively new forms of ‘interactive intelligence’ which cannot be sufficiently described as the sum of its parts (human plus tool). ‘Interactive intelligence’ can be understood as an emergent phenomenon which is described by dynamic spatio-temporal coupling between systems, embedded in a concrete social and cultural context. This conception of intelligence and intelligent behavior is opposed to

traditional accounts of intelligence where intelligence is often considered solely as a property of a system itself. A CT approach towards designing SIA also addresses the relationship between biological and artificial social agents and how such a relationship could transcend the limitation as well as hinder the cognitive and social potential of humans. Finally, the CT view on SIA is concerned with how social agent technology and human (social) cognition can co-evolve and co-adapt and result in new forms of sociality. Agent-human interaction requires a cognitive fit between SIA technology and the human-in-the-loop as designer of, user of, and participant in social interactions. Characteristics of human social intelligence like embodiment, believability, empathy, the narrative construction of social reality, autobiography and historical grounding can contribute to a cognitively well-balanced design of SIA technology, in order to improve the relationship between humans and agent tools. Cyberpets are discussed as an example of SIA technology which exemplify these concepts.

This paper tries to identify concepts which are relevant to SIA design, taking a particular human-centered stance. If the performance of an agent system, or the market success of an agent product depends on whether humans accept and enjoy using this particular tool then it is important to consider the specific way humans understand and interact with the their (social) world. Socially intelligent agents might be most successful if they are a bit like-us.

The following section identifies issential features for SIA research.

2 A brief history of autonomous agents

The terms ‘agent’ and ‘autonomous agent’ are diversely used in the literature but, as Franklin and Gaesser show, a terminology helps to clarify concepts. Their taxonomy in [FG97] subsumes biological, robotic and computational agents under the taxon ‘autonomous agents’. Is the comparison between a taxonomy of biological species and autonomous agent species purely metaphorical or are there common ‘synapomorphies’ (evolutionary novelties which originated in their closest common ancestor)?

This section gives a brief history of autonomous agents, following the definition by Franklin and Gaesser (see section 1), but focusing on social aspects which are relevant for further discussions on *socially intelligent agents*. Thus, let us for the purpose of this paper define autonomous agents as entities inhabiting our world, being able to react and interact with the *environment* they are located in and with other agents of the same and different kind. Thus, autonomous 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. What kind of agents do inhabit our world? For 3.5 billion years *biological agents* have existed whose descendants we nowadays know as plants and animals. These

agents consist of single cells, form aggregations and colonies, form complex entities by enslaving other single cells, and divide and specialize to form multi-cellular organisms. Animals and plants evolve, diversify and are able to colonize all areas on our planet by adaptation to specific biotic and abiotic constraints.

Social Intelligence. For about 2.5 million years the genus *Homo*, and for about 100,000 years the modern *Homo sapiens* species have existed on Earth. Humans turned out to be animal agents with very specific interests in other agents, in interacting, controlling, manipulating and representing them. For thousands of years they have been interested in building artifacts which imitate or depict biological agents; paintings and puppets, made of stone, clay, paper, or synthetic media, paintings or statues depicting prey, livestock, or other humans. These artifacts have been used as religious objects, luxurious gifts, efficient tools, and ordinary toys. It happened (for reasons still under discussion) that humans are above all social animals [Aro94], they survive in groups, form societies and culture, learn by tradition and education, divide labour, trade, and enjoy the company of other human beings close to them. The need to cope with complex social relationships, to acquire and manage social knowledge in order to predict the behavior of group members is, according to the *social intelligence hypothesis* ([CS92], [Byr95], [BW88]), a decisive factor in the evolution of human intelligence. The social intelligence hypothesis states that human intelligence originally evolved to solve social problems and that this capacity was only later extended to problems outside the social domain, i.e. to the domain of mathematics. Thus, mental occupation with social dynamics could have paved the way towards abstract thinking and logic. Even when primates live in a relatively predictable environment (e.g. as gorillas did before human intervention) group members are never totally predictable, they require constant monitoring, re-assessing and re-learning of relationships and group structures¹.

Anonymous versus individualized societies. Humans share complex forms of sociality with other biological agents, like social insects, species of birds like parrots and crows, and cetaceans like whales and dolphins. Natural evolution of biological social agents demonstrates two impressive alternatives of sociality, namely *anonymous* and *individualized* societies. Social insects (e.g. bees, termites, ants) are the most prominent example of anonymous societies. Group members do not recognize each other individually. 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 (see [SJA91]). The individual itself plays no crucial role, social interactions are anonymous. If we remove a single bee from a hive no search behavior is induced: Ants don’t have friends². A similar organization as in social insects can also be found in naked mole rats, an example of convergent evolution of social organization.

Primary groups and relationships. Many mammal species with long-lasting social relationships show an alternative path towards socially integrated systems.

¹I discuss the social intelligence hypothesis in more detail in [Dau95].

²Thanks to Rodney Brooks for this phrase.

Here individual recognition gives rise to complex kinds of social interaction and the development of various forms of social relationships. On the behavioral level social bonding, attachment, alliances, dynamic (not genetically determined) hierarchies, social learning etc. are visible signs of individualized societies. The evolution of language, spreading of traditions and the evolution of culture are further developments of human individualized societies. Here, primary groups, which typically consist of family members and close friends, emerged with close and often long-lasting individual contacts. A primary group is considered here to be a network of ‘conspecifics’ which the individual agent uses as a testbed and as a point of reference for his social behavior. 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. Section 3.2.2 proposes to use the term ‘autobiographic agent’ to account for this and other dynamical aspects of re-interpreting the agent’s (social) ‘history’. I suggest to generalize the concepts of primary groups and individualized societies from biological to artificial agents.

Story-telling and autobiography. Later in this paper (section 3.2.3) the importance of historical grounding of autobiographic agents for social understanding is shown. Autobiographical memory might be characteristic for humans, while generic and temporarily episodic memory might be shared with our close primate relatives and perhaps other mammals ([Nel93]). In superorganisms the colony as a whole and its creations can be considered to represent the collective memory, the history of the ‘superorganism’. With respect to social insects, evolution seem to have ‘invested’ in number, not in the complexity of the individual and its memory system. Autobiographical memory presumably defines the self (see discussion in [Nel93]). Thus, autobiographic, historically situated agents have the potential to develop a self which allows for further increase in the complexity of social relationships. As Nelson ([Nel93]) put it: “..this social function of memory underlies all of our story-telling, history-making narrative activities, and ultimately all of our accumulated knowledge systems.” Consequently, the highly individualized nature of social relationships in primates seems to correlate with the development of autobiographical memory. We might speculate that a convergent evolution took place for other mammals (e.g. whales and dolphins) and some bird species like crows and parrots, related to social living conditions and complex ways of communication. Insect and other animal species show fairly anonymous interactions because they do not have have stories to tell. Not only do they communicate through the environment (stigmergy), but the environment *is* their external memory. Ants don’t tell stories.

Communication. In human societies larger, higher level groups emerge by additional control structures. Humans seem to have an upper limit of about 150 for the size of groups with mainly face-to-face interaction and communication. According to [Dun93], 150 might, as a function of brain size, be the cognitive limit on the

number of people with whom one person can maintain stable relationships. Larger groups of people can be handled by control mechanisms like adopting roles which can be indicated by symbolic markers (uniforms, badges), or stereotypical ways of interaction (e.g. rules for greeting each other, or templates for writing and answering letters). Humans are able to handle complex social relationships by having developed very effective means for ‘social grooming’, namely the ability to communicate by an elaborated and efficient communication system, language, which allows communication about issues on different levels of abstraction but is less immediate than communication by ‘body language’ and facial expressions. Modern humans use language primarily to communicate stories about other persons which indicates how closely language is related to this primary function.

“In human conversations about 60% of time is spent gossiping about relationships and personal experiences...The acquisition and exchange of information about social relationships is clearly a fundamental part of human conversation. The implication, I suggest, is that this was the function for which language evolved.” (R. Dunbar, [Dun93])

Believability. 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 behavior ([PP95], [Den87]), indicating a bias to perceive and recognize other biological agents. Humans have a natural tendency to animate and anthropomorphize nature ([Wat95]). Humans are not the only tool-designers in the animal world, but they happen to be the best ones, in terms of creativity of using material and functionality of the results. For a few years humans have been developing specific agents based in silicon. Part of these *artificial agents* is made of software. *Computational agents* which can take on different forms are called ‘mobile agents’ when navigating networks, but called ‘intelligent agents’ when they solve tasks which humans did before. They assist 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 extensively 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* [Bat94] give an ‘illusion of life’. They need not necessarily appear or act just like biological agents, but some aspect in their behavior has to be natural, appealing, life-like. Research in believable agents benefits significantly from animation work and artistic skills to create fictional, imaginary but believable creatures. Section 4.1 goes into more detail.

A parallel development has occurred in the development of mobile physical tools, *robots*. For decades ‘life-like’ robots had only been known from science fiction

literature. Robotic systems existed in production lines or other areas outside usual human experience. 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 behavior has been studied with these physical artificial agents, namely robotic agents. Since humans tend so naturally to bond to biological agents, their artificial counterparts, too, will 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 over their roles.

Thus, research on computational and robotic agents has steadily converged towards common issues in a domain where an important part of the functionality of artificial agents is interaction with humans. Issues of agency, believability and sociality are examples for commonly arising research issues. Mechanisms like perception and communication are central. Thus, learning about the 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 characterized by an overlapping of the domains of biological, computational and robotic agents that the question arises whether a common ‘social interface’ should be considered, either as a conceptual construct or a technical implementation.

Cross-species interactions. Is it possible and desirable to construct ‘social interfaces’ as technical or conceptual spaces in which different ‘species’ of agents can become engaged? Software agents and physical agents (robots) need not necessarily have a ‘natural’ form of social behavior, communication and interaction. They might 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 subgroups, ‘dialects’ of communication and interaction within larger groups, etc.), influenced by domain-specific requirements and constraints. Specific dynamics and expressions 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. Extensive research is currently studying ‘cross-species’, ‘natural’ forms of interaction between humans and robotic or software agents, e.g. [KDK97], [RS97]), [WR97]. Terms like ‘human-robot symbiosis’ ([KPI96], [WPAK97]), ‘mixed-initiative problem solving’ ([Tat97]), or ‘co-habited mixed realities’ ([vdV97], [CvdV97]) are examples of research activities in this field.

3 Human social intelligence

Social issues are studied in various research fields like psychology, sociology, ethology, economics and others, and this paper does not try to achieve a universal definition of what social intelligence is. Different research topics can be studied like: comparative studies of different forms of sociality in animal species; intelligence as a cognitive capacity of an individual versus distributed, shared models of the emergence of intelligence; social intelligence as a byproduct of general intelligence, or as one of several faculties of intelligence versus social intelligence as the primary form of intelligence; phylogeny of intelligence in primate societies; learning versus innateness models of the ontogeny of intelligence; the ‘purpose’ of social intelligence as a prerequisite for complex forms of cooperation and the division of labour, etc. For our SIA considerations in this paper I focus on the following aspects of human social intelligence: humans as embodied, empathic, autobiographic, narrative agents.

3.1 Embodiment

Embodiment is naturally given in biological agents but quite difficult to define for artificial agents. It is a trivial statement to biologists that all biological systems have a body, that they are living through their body, their existence cannot be separated from it. The issue of embodiment has recently attracted particular attention. Opposed to traditional 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’ ([Pre97]). EAI stressed the need to study intelligence in an embodied system. The emphasis on physical agents led to cognitive robotics [Bro96]. Recently, discussions have started on what embodiment can mean to a software agent [Kus97].

The issue *that* embodiment matters for intelligence, life and agency is nowadays widely accepted. But the question of how and to what extent embodiment matters are 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 tiny memory space inside a computer?³ Can inputs (e.g. keyboard commands) and actions (e.g. UNIX commands) really be considered comparable analogues 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 on finding the ‘right’ levels of comparison is still open, yet the danger of ending up in frameworks based on pure metaphorical comparisons is obvious. However, the observation that robotic and computational agents can *appear* life-like and are often described and treated as ‘personalities’ ([TP97]) or ‘characters’ indicates a human tendency to ‘animate’ the world and even by itself justifies the attempt to discuss concepts of human (social) intelligence in the realm

³See discussions by Tom Ray on artificial ecosystems, [Ray92], [Ray94].

of SIA. The better artificial robotic or computational agents can meet our human cognitive and social needs, i.e. the more they appear ‘like us’, the more familiar and natural they are and the more effectively they can be used as tools.

The next sections discuss the specific way humans construct, understand and interpret the social world.

3.2 Understanding social agents

In the following I discuss the concept of ‘stories’ which humans create to maintain a concept of self and to communicate and understand social interactions.

3.2.1 Memory and Stories

In AI and more generally in computer science the concept of memory was dominated by the technology of digital computers. The main metaphor was to consider a memory system as a huge data-base where static ‘memory items’ (information about objects, situations, rules, abstract knowledge) are stored away and, at a later stage, identically retrieved. This metaphor has also had a strong influence on concepts of human memory in cognitive science⁴. Recently, story-telling systems have increasingly been studied in AI, e.g. see [Dav96], [Sha97],[Mat97], [HRvG97]. Such systems can make exciting entertainment products, but their significance goes beyond that, namely they can indicate a paradigm shift in AI: Increasing evidence in psychology shows that human understanding and interpretation of the world, in particular the social world, is based on stories. The construction of reality, the organization of remembering, dialogue, and social interaction seem to be grounded in narrativity. According to the psychologist Jerome Bruner, narrative seems to be the form by which we not only represent but also constitute reality ([Bru91]).

In [Wye95] Roger C. Schank and Robert P. Abelson give an argument for the relation of stories to knowledge and memory and the role of stories in individual and social understanding processes. Based on their work on *scripts* as representations of generic event memory, e.g. a prescription of how to behave in a restaurant ([SA77]), they propose scripts as a suitable computational approach towards building story-telling systems. They hypothesize 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. Dialogue can then be understood as the production and re-construction of stories which are most similar to the ones which are produced by the dialogue partner. Such a dynamic account of human memory goes back to work done by Bartlett more than

⁴In [DC96] this issue is discussed in more detail.

half a century ago ([Bar32]). A social origin for story-based human memory and understanding is hypothesized by Read and Miller. In [RM95] they address the evolutionary significance of stories and assume that social living conditions might have favoured ‘naturally’ the evolution of story-telling mechanisms in human cognition, since stories seem to be efficient means for managing social interactions.

“Stories may be the only possible way to deal with the enormous complexity of human social interaction..., it is because of the social, and the need to effectively manage social interactions, that we developed stories... It is our stories that make us human.” [RM95], pp 148–150.

Robert Worden uses in [Wor96] scripts in order to model primate social intelligence. He proposes a working computational theory of primate social intelligence. He uses ‘scripts’ (consisting of mental models, production rules and *scripts* as defined in [SA77]) as representations and defines computational operations on scripts which are, in his view, sufficient to support social learning, planning and prediction. He compared his model with primate data and found good correlations. It seems to be a very interesting approach towards modelling and describing primate social behaviour, and an excellent tool to evaluate and discuss ethological data. However, as psychologists point out (see [Bru91], [Nel93]), Schank and Abelson’s scripts bear the problem that they only capture generic, canonical behavior in a culturally defined situation. But a story becomes worth telling by breaches and violations, by individual properties ([Bru91]). Thus, scripts are abstract data-structures which can represent and guide repetitive behavior, they abstract away from the individual, the embodied agent, who is telling his/her stories, relating to own experiences, constituting the autobiography. “Autobiographical memory forms one’s personal life history” ([Nel93], p. 8).

3.2.2 The Autobiographic Agent

In order to account for the life-long dimension of human re-construction of the own history and personality, I define in [Dau96] an *autobiographic agent* as an embodied agent which dynamically reconstructs its individual ‘history’ (autobiography) during its life-time. Autobiographical memories are widely studied in psychology (e.g. [Con96a], [Nel93]). A constructivist, dynamic account of remembering suggests that “memory is primarily a vehicle for personal meanings and for grounding of the self, and that accuracy is secondary to this role” ([Con96b]).

An important aspect in AI research on knowledge and memory is *consistency*. Various algorithms have been developed in order to build up and manage a ‘complete’ and consistent knowledge or database. On the other hand, humans easily seem to cope with this problem. But there is much evidence that the problem of consistency itself is an artificial one. Instead, the subjective impression of being a static ‘personality’ is an illusion and might only be a good approximation on small time-scales ([Bar32]). Humans seem to integrate and interpret new experiences on the basis of previous ones. Previous experiences are reconstructed with the actual

body and concrete context as the point of reference. In this way past and presence are closely coupled. In combination with human capabilities of rehearsal (as the basis for acting and planning) this coupling is linked to the future ([DC96]). Humans do not seem to worry much about consistency, they give explanations for their behavior on the basis of a story, a dynamically updated and rewritten script, their *autobiography*. Believability (see section 4.1) of this story (to both oneself and others) seems to be more crucial than consistency. This is what characterizes an autobiographic agent. A CT approach to SIA technology has to take into account that humans are autobiographic agents, that they interpret interactions with reference to their ‘history’ and bodily grounding in the world.

The behavior 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). These ideas on historical embeddedness of humans and other animals are in line with Hendriks-Jansen’s work which gives in [HJ96] a strong argument for the importance of situated activity, interactive emergence and the ‘history of use’. Thus, social behavior can 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 humans are constantly telling and re-telling stories about themselves and others. Humans are *autobiographic agents*.

3.2.3 Social understanding: Stories about oneself and others

Is human social understanding basically computational, i.e. is it about matching of scripts and stories about others, manipulating symbols, data structures and representations? An alternative, phenomenological view is suggested in [Dau97] where I discuss that social understanding emerges from internal dynamics inside an embodied system. Social understanding, as a form of ‘communication’ is based on empathy as an experiential, bodily phenomenon of internal dynamics, and on a second process, the biographic re-construction which enables the empathizing 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 is supposed to be a necessary condition for synchronized coordination processes (e.g. in verbal and non-verbal communication), and a prerequisite of ‘true’ social understanding, as opposed to models of social understanding on the level of data structures.

Biographic re-construction as a crucial mechanism in human social understanding is based on the re-construction of a biographical ‘story’ about another person. Elaborated, typically human kinds of empathic understanding of another person can

be thought of as creating a plausible story about the person’s context, the biography, including aspects of past, present and future. This creative aspect of story-telling, i.e. to tell autobiographic stories about oneself and biographic re-constructions about other persons, is linked to the empathic, experiential way of relating other persons to oneself. I hypothesize that this is the central set of mechanisms which constitutes what we call ‘social intelligence’.

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, fairy-tales, 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. In section 4.2 we make the connection between stories and the actors enacting the stories (e.g. virtual pets).

4 Agent technology from the observer point of view

This section argues for a balanced, historically grounded, socially situated, and ecologically plausible design philosophy of believable (social) agents.

4.1 Believability

The concept of ‘believable interactive characters’ originated from arts and was introduced by Joseph Bates for software agents ([Bat94]). The concept has resulted in believable artificial software characters and personalities (e.g. [LB97], [Rei97], [TP97]). Believability has also been discussed for autonomous robots, e.g. in [Dau97]. The key aspect is that believability does not necessarily depend on intelligent, complex or realistic behavior, believable agents need not show ‘intelligence’. ‘Believability’ is in the eye of the observer which means that it is influenced by the observer’s individual personality, naive psychology and empathy mechanisms ([Dau97]). Thus, whether a specific person finds an artifact and its behavior believable or not depends on his/her own subjective perception and interpretation of the artifact and the context the artifact is behaving in, as well as on the social and cultural context which the human is living in. Building believable artifacts can therefore hardly be guided by ‘objective performance parameters’. Good examples of believable characters are Toy Story or Luxo Jr. (both by John Lasseter, Pixar Animation Studios). As I discussed in [Dau97] humans are biased to interpret the world in terms of intentionality and explanation. Humans seem to be automatically inclined to judge any artifact according its believability. There is however a significant difference between Luxo Jr. and Toy Story which are mentioned above. Luxo Jr. is a computer animated

story about parent and child desk lamps. They do not mimick the form and shape of any human or animal (unlike the animated human-like puppets which act in Toy Story as the main characters.). The desk lamps look ‘alive’ because they show behaviors which are typical of animals: giving attention, playing, social behavior etc. These are all ‘entry points’ which allow the observer to match the artifact’s behavior with behavior which is shown by living systems. However, the lamps do not mimic the morphology of any specific animal or any specific species. In this way, they demonstrate clearly that even systems which are inherently different from natural living systems can show ‘life-like’ properties, so that humans find them engaging, appealing, and immediately attribute intentionality, mental and emotional states. Thus, believability of technology should not be considered simply an add-on to make existing products more appealing or ‘cute’, e.g. true believability is not the idea of attaching a tail and big eyes to back and front end of a robot. The latter would be an example of a ‘shallow’ approach to believability. Taking believability seriously directly points at typically human ways of perceiving and interpreting the world. Believable technology is ‘familiar’ to humans, it meets their cognitive and social, typically human needs.

The ways people react to believable agents point towards 1) the social and emotional dimension of computer technology and 2) in this way, a challenge to traditional conceptions of intelligence and the design of intelligent systems. A software engineering process of building a piece of software does usually not consider what kind of emotions humans might project onto the product. This aspect of the ‘human-in-the-loop’ is historically rooted in second-order cybernetics (Heinz von Foerster) which studies systems involving the observer as a constitutive part of the process of knowledge creation. Transcending objectivity in building systems has therefore a long tradition, and believable social agents can possibly put light on how such systems can be designed. Phoebe Sengers discusses in [Sen97] that builders need to have tools which enable them to build agents whose goals and intentions are communicated/signaled clearly and effectively to the audience. In her view the process of building social agents has to become social as well. Thus, properly designing SIA is not at all a trivial task. Nevertheless, believable agents are sometimes said to be scientifically ‘cheating’, since they put all the intelligence in the human-agent interface and rely on the intelligence of the human using and interpreting this interface. This is true, but is it a bad point? It is only a bad point if the goal of scientific research on agents is assumed to put intelligence into the agents themselves, a traditional AI attitude which has been overcome by the recent paradigm shift from algorithms to interaction ([Weg97]). Notions of ‘interactive intelligence’ have a different underlying ‘philosophy’, which is no more or less scientifically valid than the traditional AI approach.

The concrete technical basis of SIA technology (e.g. whether software or hardware) does not seem to matter much. Humans are from their early childhood on experts at taking various abstract or fictional things for real entities (comprising comic, television or video game characters, football teams as well as political theories and religion). What seems to count is the question of what is real to the embodied

mind of an individual person. Following the argumentation of radical constructivism (e.g. [Rot94]), it is more useful to discuss the individual's constructed conception of reality, the *Wirklichkeit* than an objective reality. The meaning, and not the technological basis, is central. In this way, experiences which are important to the life of an individual should be taken seriously, no matter if they originate in interactions with real, simulated, virtual or fictional entities. In discussing a CT approach to SIA technology I therefore do not distinguish between technology for constructing robots and developing computer programs, or even writing novels or science fiction stories with believable characters. Knowing more about the co-adaptation of technology to human cognition and the social context, and the way how humans in reverse understand, interpret, and interact with technology can result in believable, interesting products.

4.2 Believable virtual agents: virtual pets

Software or virtual pets are the latest development of believable product technology emerging from a cross fertilization of artificial life and software agent technology. The resulting offspring known as 'cyberpets' are popular applications of artificial life and artificial intelligence (agent) technology, e.g. Creatures (Cyberlife, [GCM97]), Petz & Dogz (P.F. Magic, [FSR97]), Fin Fin (Fujitsu), Tamagotchi (Bandai). These cyberpets are 'life-like' not necessarily with respect to their appearance, but by the fact that they are living in an environment, can express emotions, can die (therefore have a 'life-time'), and last but not least, can interact with a human user. They are in a virtual sense fairly 'complete organisms' which are generally embedded in a more or less complex 'story'.

Research in believable agents has demonstrated the central role of the human designer of, user of, and observer of agents. Believable agents interact 'naturally' with their users, they appear 'life-like'. Users are inclined to become emotionally bonded to believable agents, and in the case of virtual pets it can develop to the extent that humans adapt their daily routine to cyberpet welfare concerns⁵. As already mentioned, criticisms have therefore come up that such products are 'cheating', i.e. pretending to be more interesting or 'intelligent' than they actually are, or that they 'exploit' natural human instincts of nurturing and caring. Believable agent research has been strongly influenced by animation technology (Disney's 'The illusion of life', [TJ81]), and indeed animations can 'cheat' in the sense that they can present an implausible, unrealistic 'fictional' reality, e.g. computer animations of dinosaurs or animated comic characters (another example of believable characters is given in the next section). Such a perspective widely assumes a 'passive' viewer, the recipient of the presented story. The stories are generated by the designers, and assimilated by the viewers. Each viewer can have a slightly different interpretation and associations with the story, but the basic script of the story is socially shared by a large group of viewers. Current research, e.g. by Glorianna Davenport's group

⁵A whole (parallel?) world has developed around cyberpets, e.g. vets and cemeteries for departed pets.

(Interactive Cinema Group, MIT Media Lab, Massachusetts), tackles the issue of developing story-telling, interactive media ([Dav97], [Dav96]), and making cinema more active and interactive, but there is still a long way to go until this becomes familiar technology.

The still passive role of today's animation and movie viewers is different from written material like books, a medium which can in the form of a novel or fairy-tale also tell a story, however, the reader is required to reconstruct the story in much more detail. Not only to fill in details (concrete shapes, forms, faces), but also to actively fill in gaps by means of its own imagination and personal experiences. In contrast, interactive media like cyberpets have the potential to create socially shared, and individually experienced stories at the same time. The default setting e.g. the ecosystem where the Norns (in the computer game *Creatures*) are living is the same for all products, but the agents develop over time, they learn and adapt to the user's behaviours and reactions to them. In this way the agents become unique, develop an individual biography, a life-time story. Thus, being embedded in a believable (interactive) story makes agents believable. The 'story' does not only comprise the agent itself, its behaviour and appearance, but also its environment including other agents, and humans.

We discussed above the importance of individuality and history (autobiography) for natural living creatures and suggested that these concepts can also be used to enhance the believability of artificial creatures. Such interactive, socially and historically embedded artificial life-forms are not simple results of anthropomorphic projections, when the argument of 'cheating' might be justified. Cheating becomes visible in situations when the complexity of the agents appearance does not match its behavioural and interactive potential. An unbalanced design might result in a mismatch of user's expectations and agent's performance, resulting in a raising frustration level of the user. A major theme in biology, 'form fits function' captures this points, and nature gives examples of balanced designs where behaviour and morphology of animals fit their functions very well (e.g. the evolutionary 'remodeling' of tetrapod forelimbs to flying, swimming, climbing, running). An agent who appears humanoid is supposed to behave human-like. If it does not, then users might become disappointed since wrong expectations were created.

Cyberpets are not 'complex' (or 'intelligent') in themselves. What makes them special is the fact that they exhibit interesting behaviours only in the interaction space of agent and user, i.e. only over the course of the interaction between agent and user. Social bonding cannot be generated by the agent, or the user alone. But by agent and user interacting with each other, new forms of interesting behaviours on a different level of complexity can emerge. *Interactivity* is the key point which makes cyberpets so believable and popular, and can even compensate for simple designs: Key chain product (like Tamagotchi from Bandai or Tiger Electronics's Giga Pets) are very popular, although they use simple technology and the agents show a poor degree of individuality, expressions of life, or means of interaction (e.g. only a few buttons are used to give a Tamagotchi feedback from the user). Interactivity can provide the illusion to be treated individually, even if the cyberpets do only react

to certain standardized input from the user. Thus, whether we ‘like’ them or not, ‘life-like’ artificial agents, e.g. cyberpets, are examples of how humans view and interact with the (social) world, how they are biased to interpret the world in terms of intentionality, and how much humans need the feeling of ‘belonging’, and to be engaged with the world.

4.3 Believable Fictional Biology: The Case of Rhinogradentia

In 1961 the first edition of Harald Stümpke’s (alias G. Steiner) monograph ‘Bau und Leben der Rhinogradentia’ was published in Germany ([St9]). The book was later translated e.g. to English (title ‘The Snouters’) and French and appeared in many editions. In 1963 it was reviewed by G. G. Simpson in the international journal *Science* [Sim63].

‘Rhinogradentia’ is the name for an order of mammals which is divided into 14 families and 189 species. They were said to have been discovered in 1941 and are endemic to a group of islands in the Southsea called Hi-lay. The monograph gives a comprehensive introduction to the habitat, evolutionary aspects (adaptive radiation), taxonomy, morphology, ontogeny, and various other aspects of living and surviving of the snouters. Plates 1 and 2 show an excerpt from the variety of species and biological issues presented in the book⁶. The existence of these isolated group of species had been ingeniously foreseen by Christian Morgenstern whose German original of the following poem (‘Das Nasobem’) was published in 1905.

Along on its probosces
there goes the nasobame
accompanied by its young one.
It is not found in the Brehm,
It is not found in Meyer,
Nor in the Brockhaus anywhere.
‘Twas only through my lyre
we know it had been there.
Thenceforth on its probosces
(above I’ve said the same)
accompanied by its offspring
there goes the nasobame.
(Translated by L. Chadwick)

The scientifically presented story about the ‘snouters’ is fictional, although, as described by Karl D. S. Geeste in [Gee88], numerous readers mistook the story for real. The story is presented in bookform, and there is no direct interactivity with the user (reader) as it is possible with today’s technology of Cyberpets (see previous section). However, a fairy-tale or a novel is usually not reviewed on two pages in an international scientific journal, as was the case here. So, what made this case special,

⁶Drawings are included with copyright permission by Gustav Fischer Verlag.

why is the story of the the ‘snouters’ attractive, and believable to both biologists and laymen?

‘Snouters’ are different from dragons or other legendary creatures which might resemble animals, or rather chimerae, which are not viable forms of life. Dragons could not exist on Earth, if they could exist at all. As Geeste points out there are some details of biological implausibility in the monograph of the ‘Rhinogradentia’, but in general they could possibly live, e.g. they are rather alternative life-forms than pure products of unconstraint fantasy. The ‘snouters’ have body plans which clearly identify them as mammals. Thus, the ‘snouters’ represent a fictional artificial life form, life-as-it-could-be, a possible result of conditions when evolution favours changes in functionality of a mammal’s nose. In contrast to Cyberpets and animation characters the ‘snouters’ also have a phylogenetical and clear ontogenetical history⁷. The reader gets an idea how the ‘snouters’ the might have developed. The ‘snouters’ share with many cyberpets the fact or fiction that they are living in an environment (cyberpets usually have a house where they sleep etc.). However, environment has a much richer meaning for the ‘snouters’ since they are adapted to the environment and embedded in an habitat, an ecosystem involving other species members, prey and predators, etc. Thus, it makes them believable that we know about the conditions of the environment where they evolved and where they live now, which makes their appearance and the description of their behavior much less arbitrary.

For the development of agents in general, what can the ‘snouters’ teach us? 1) They could be alive! Life-like artifacts need not necessarily mimic nature, but it helps if their appearance and behavior are biologically plausible. It is hard to imagine an alternative biology (e.g. ‘evolution’ without genetic material), but based on what we know about biology on Earth we can think of alternative life-forms. This also includes extinct species, e.g. animals of dinosaurs usually appear plausible. 2) ‘Form fits function’ and ‘function fits environment’ can make a balanced, ecologically plausible design. Thus, social agents could become more believable when they are embedded in a rich, and plausible story. The more we know about the life and ‘autobiography’ of our agents, the more believable they can become.

5 Social Agents within an Embodied Artificial Life framework

This section attempts to relate the process of designing believable agents to a research methodology, namely Embodied Artificial Life. As the previous sections pointed out, building believable and social agents is not only interesting in terms of the end products, it addresses many exciting issues concerning (human) intelligence and ‘sociality’ in general. Thus, a CT approach towards SIA comprises two dimensions, product design on the one hand and an agenda for basic research on the other

⁷Cyberpets show an ‘ontogeny’ by developing into different characters, a simple simulation to demonstrate the basic idea of ‘growing up’, to have a biographic history.

hand.

5.1 Sciences of the Artificial

This section discusses methodological issues, namely how and why social agent design (and the concepts discussed so far) can be interpreted along the Embodied Artificial Life (EAL) direction, a bottom-up approach towards (social) intelligence.

EAL is a specific research direction which is rooted in Artificial Life (AL), but differs significantly in its methodological approach. AL is fundamentally different from an older research field, namely Artificial Intelligence (AI), which is also concerned with the construction of artifacts. However, the psychological concept underlying AI is the Physical Symbol System Hypothesis (PSSH), the assumption that human intelligence is (or in its weaker formulation: can be modelled as) computation. Research in AI has long been focused on human problem solving, i.e. the complexity of human life and living was somehow reduced to problem-solving. Aspects of development and evolution (ontogeny and phylogeny), embodiment, personality, individuality, creativity, social/emotional skills have long been regarded as non-significant.

This has had also strong influences on the concept of intelligence, which has been considered as a numerical quantity (or a set of quantities) which can be measured by tests on logic, mathematical skills, spatial thinking, planning, linguistic abilities. These quantities are not only used for scientific purposes, they have shaped public opinion about the ‘utility’ of a person, about the position of the human species in nature, about the human society (especially the educational system), and about the attitude of humans towards other animals. This conception of (human) intelligence has been applied to other animals. Intelligence tests comparable to the ones used for humans have been used to assess under laboratory conditions intelligent skills of our closest relatives, apes, and also other animals. The very idea that a number can be calculated which reflects the ‘value’ of a living being allows one to make comparisons between members of the same species as well as among species. On the other hand, all that we know about nature, biology, and evolution does not give us a single clue that there exists any plausible dimension along which living systems should be ‘evaluated’. It is inherent in the very idea of evolution by natural selection that species living at a certain period of time are equally ‘successful’. Any given species is not ‘better’ than another.

The basic characteristics of AL research⁸ can be summarised as follows:

- Systems are not to be understood by analysis and decomposition, but by synthesis (bottom-up approach).
- Based on local, non-linear interactions between components, the emergence of complexity on the next higher level is studied.

⁸For a comprehensive introduction to AL see [Lan89].

- For non-trivial systems emergence cannot be predicted, neither can a system be designed with a particular pattern as a result. Consequently, we either have to put the system in its concrete environment and let it ‘run’, observing the outcome, or we can apply evolutionary techniques in order to adapt the system to its environment.

Fig. 1 shows two AL research directions. Fig. 1 a) shows the probably most widespread direction, driven by the underlying assumption that objective criteria exist to develop a ‘logic of life’ which is independent of the matter in which the life-form is realized. This ‘logic of life’ should be the language which produced biological life, which can produce artificial life (robots, software creatures) and which potentially also applies to other, unknown, yet to be discovered alien life forms on other planets. This direction is most closely related to artificial intelligence research. The techniques which are used might belong to artificial life research (e.g. bottom-up approaches to program and design robots), nevertheless the underlying ‘philosophy of thinking’ is still rooted in AI. Along this direction the issue of ‘objective’ criteria for life are discussed, and how tests (analogous to the Turing Test for human intelligence) can separate living from non-living artifacts. As it is described in [Lev92] the issue of finding criteria for life was under discussion right from the beginning of AL research (and has not yet been solved in biology). An agreement on life criteria is seen as crucial, in the same way as artificial intelligence research has always been based on conceptions of intelligence. The method of using biological systems as models for artificial systems is also strongly related to the assumption of being able to identify a specification of life and mechanisms of ‘living’ which apply across species and across the physical basis.

Fig. 1 b) sketches an alternative approach, the creative ‘story-telling’ approach. The goal is to *find forms of complexity in artificial media which appear to be natural, and which give us plausible explanations about life*. Here, we need not assume the existence of a ‘logic of life’. The challenge is rather to explore and investigate what form complexity can adopt in artificial media, how it can evolve and/or how it can be designed. We need not only study evolution in order to investigate complexity. Design of complex systems can give us plausible explanations as well. Each material can have its own inherent ‘logic’, i.e. properties and mechanisms which form complex structures. In addition, why is it necessary to define life? Life (and intelligence) are in the eye of the beholder (i.e. in the mind of the observer), these concepts characterise the way human beings interpret the world. Why should scientists by all means define life, if every individual human being has his/her own conceptions of life and living? Nevertheless, the research agenda is clearly defined, it is about investigating complexity, finding believable ‘scientific stories’ which can give us plausible explanations and interpretations of life.

The difference between a) and b) is not identical to the weak/strong interpretation of AL, comparable to the distinction in AI: *Is our particular artifact intelligent/living or does it only behave like this*. Moreover, the difference between a) and b) does not depend on the use of particular architectures or mechanisms. The crucial

difference is rather whether one believes that something is ‘out there’ which has to be discovered (the logic behind the things, the logic of life), or whether it is the very process of creating, constructing, exploring, designing systems which helps us to learn about life. This is my answer to Simon Penny’s question posed in [Pen95]: “Why do we want our machines to seem alive?” Artificial agents can help us to learn about life!

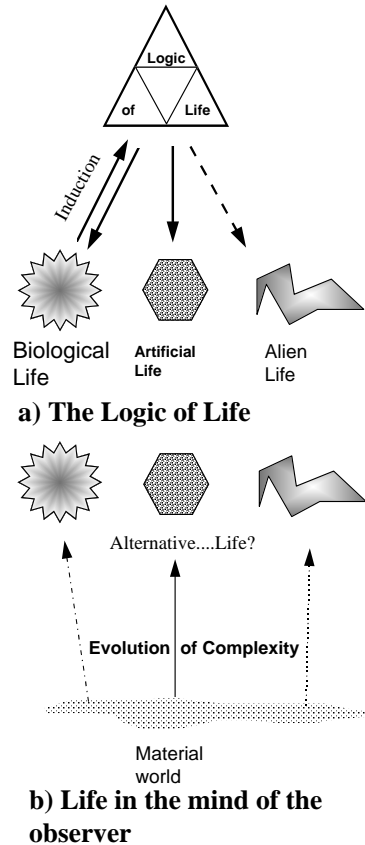


Figure 1: Two different frameworks of artificial life.

The term *Embodied AL* (EAL) is generally used for the specific research direction which is sketched in fig. 1 b), namely an approach which is more concerned with the specific instances of life, individual properties of the matter and its internal dynamics, and less concerned with the search for a universal logic of life and generic mechanism which can implement this logic. AL was originally based on the synthesis (as opposed to analysis) of systems. However, recent work tends to investigate

generic mechanisms, formal ‘languages’. These models or descriptions are sometimes used in a second stage as implementation guidelines, abstracting away from the specific properties and ‘history’ of the system, or feeding in these parameters as ‘implementation details’. The emergence of complexity is therefore decoupled from the physical basis. Abstractions are useful for analyzing and understanding a system but when building an agent its actions have to be grounded in the real world. AI robots, equipped with a navigation program and a complete map of the environment have often failed to run robustly and ‘intelligently’. Then, behavior-oriented and ‘New AI’, focusing on sensori-motor couplings and behaviors exhibited in a concrete environment has outlined an alternative direction [Bro91b], [Pfe95]. Recent discussions on embodiment and AI indicate that there is more to an intelligent robot than simply putting a computer on wheels. AL seems to be in danger to end up in the same trap, by assuming that we can make a system alive by studying logic and implementing it on computer or robot hardware. Therefore Embodied AL is an alternative approach towards life-like artifacts which originated in work e.g. described in [Mae90] and [SB95]. The EAL framework towards intelligence and ‘aliveness’ focuses on the grounding of these conceptions in the concrete embodied, situated system, and studies how (from an observer point of view) interactions with the environment can bias the attribution of intelligence and ‘liveliness’. EAL takes into consideration the specific physical properties of matter, grounding intelligence/life in the social and cultural context. The question of whether or not these systems ‘are’ intelligent/alive or not is not relevant. The more cognitive oriented concepts like believability and human remembering processes which are discussed above can easily be interpreted in terms of EAL, because they are addressing the dynamics inside an agent and its coupling to the environment. On a different systems level Tom Ray’s research ([Ray94]) which is also about finding the natural form of complexity in artificial media is related to the EAL approach. However, the previous section of this paper discussed systems and concepts on the basis of already very complex ‘media’, namely human beings and society, and not on the level of molecules or single cells (what is most common in AL) or pieces of program code, see Tom Ray’s *Tierra*, a digital ecosystem.

A potential misinterpretation of EAL when it comprises concepts like ‘believability’, ‘embodiment’, ‘autobiographical agents’ and ‘social intelligence’, is the argument that this is an attempt to ‘subjectivize’ science and that it would imply a departure from scientific experimentation and scientific methodologies, as a ‘post-modernist desire to dissolve everything into mush as a dead end’⁹. However, using subjective concepts and denying an ‘objective reality’ does not mean that scientific experiments and controlled reproducible experiments and testing of hypotheses are no longer relevant. But EAL provides a particular viewpoint, points towards particular research issues, and implies a different way of evaluating systems which are built for humans.

Thus, Embodied Artificial Life can provide a framework for socially intelligent

⁹Thanks to an anonymous referee for ECAL97 for expressing this opinion.

agent research, making the link between scientific methodology and experimentation and subjective, ‘artistic’ ways of creating artifacts.

5.2 About (Not) Modelling the Social World

Section 3.2.3 discussed an approach towards human social understanding based on the analysis of ‘internal dynamics’. Two mechanisms were proposed to play a role in human social understanding: (A) Empathic resonance, (B) Biographic reconstruction. Both concepts describe how the single agent makes contact with others and relates to the (social) world. They bridge the gap from the inside, the internal dynamics occurring in a single agent, to the external (social) world. And they synchronize the dynamics which can arise from interaction between two or more ‘social’ agents. Although these concepts might appear fairly ‘abstract’ to some readers, I discuss in the following that they are in line with the behaviour-oriented, bottom-up paradigm towards artificial intelligence.

It has been a long tradition in artificial intelligence research to explicitly model social expertise, e.g. the area of BDI (Belief/Desire/Intentions, e.g. [KGR96]) is taking the ‘intentional stance’ of studying architectures and theories where agents are attributing beliefs, desires, wants, abilities and other mental states to other agents. According to Dennett ([Den87]), the behavior of an intentional system can be predicted by the attribution of mental states. The symbolic representation and computational manipulation of beliefs, desires, etc. play a strong role in current intelligent agent research, see [WJ95]. However, having found the descriptive account of interaction does not answer the question how sociality can be grounded in real world interactions of embodied and social systems. The argument that humans ‘naturally’ attribute intentions and other mental states to an agent does not necessarily imply a symbolic approach towards building agents.

This debate seems to parallel discussions on how to make robots intelligent. In [Bro96], Rodney Brooks gives a ‘historical’ overview on the shift of viewpoint from classical to behaviour-based robotics. Brooks was one of the strongest proponents of this shift of viewpoint or paradigm. Nowadays the traditional AI and ‘nouvelle AI’ direction co-exist, the question of which direction is building ‘better’ systems is still open, partly due to the fact that the evaluation of the *behaviour* of a system is often less straightforward than measuring its performance in solving a particular *task*. A key point in Rodney Brooks’s argumentation for behaviour-oriented robotics was situatedness, i.e. the robot interacts in the real world, completely depending on on-line, real-world sensor data which are used directly in non-hierarchical behaviour-oriented control architecture (e.g. the subsumption architecture). The alternative, ‘traditional’ function of sensor data is to use them as input to build up and update an internal world model, where relevant features of the environment (e.g. geometrical information) are represented, and reasoning mechanisms are making inferences on this representation. In contrast, according to the situated stance, “The world is its own best model” ([Bro91a]). In other words ‘being in the world’ provides context, sensory feedback, and the ‘right’ perspective. In this way information is grounded

in ‘meaningful’ experiences which trigger appropriate reactions.

The behaviour-oriented approach towards robotics (and towards intelligent systems in general) has a flavour of behaviourism, a particular school and way of thinking and experimentation which had dominated animal behaviour research for decades ([Wat24], [Ski74]). Behaviourism considers mental, and in general internal states as irrelevant and focuses on the external, ‘objective’, measurable behaviour of an animal. However, in my view there is no need to identify behaviour-oriented AI with behaviourism. Behavior-oriented AI does not question the existence or relevance of internal dynamics or psychological processes, it only rejects the idea that these are based on symbolic, internal world models as they have been used in traditional AI.

However, many behaviour-oriented approaches toward robot sociality focus on stigmergic communication (see [Mat95] for an overview on collective robotics research) which means indirect communication via interaction with the environment, a more external conception of sociality, based on anonymous interactions. Such social behaviour is characteristic of anonymous social insect societies. On the other hand, for individualized societies (primates, whales, dolphins), social interaction is basically individual, personal, and internal dynamics and emotions do play an important role. Individual, internal dynamics can best be experienced when one is directly immersed in a situation. Personal face-to-face communication is by most people still preferred to other indirect forms of communication, especially in those situations when the outcome of a meetings depends on picking up the right social cues about motivations and intentions of other persons. Being immersed in a social situation makes it much easier to pick up relevant information than e.g. by email, telephone or even video-conferencing, although ‘rational aspects’ of communication (e.g. providing facts, tabular data) can be done through these media equally well (or even better). A robot can best ‘decide’ on how to avoid a wall when it is facing it and getting concrete sensor data. The best way for a human being to behave socially results from being situated in the real social world. For an artefact which should behave in a human-like style of social intelligence it might be helpful not to assume that beliefs, desires and intentions have to be modelled explicitly. Social understanding could rather be grounded in experiential processes of internal dynamics, which self-organize from physical presence, situated in a social situation. To conclude:

The social world is its own best model.

In [HJ96] Horst Hendriks-Jansen gives an excellent example of how social interaction, namely turn-taking between a mother and her baby, emerges without any mechanism which is explicitly controlling turn-taking: a mother responds to her baby’s pauses in sucking with jiggling in order to encourage the infant to resume sucking. The success of this emergent turn-taking (jiggling, sucking) relies fully on the mother’s interpretation of the baby’s behaviour, although the mother received a message that was never sent. Biologically, the infant need no encouragement to resume sucking. Sucking is rhythmical, but the duration of bursts and pauses is

random. Actually, as Hendriks-Jansen points out, jiggling reduces the likelihood of the beginning of a new burst. The cessation of jiggling encourages the baby to resume sucking.

Another example from the area of cognitive robotics was described by Rodney Brooks ([Bro97]). Again the subject is turn-taking behaviour, in this case between a humanoid robot (Cog) and the human experimenter. From an observer point of view they are playing with a ball: Cog and the human reaching out, grasping, and putting down a ball in alternation. The behaviour of the robot is based on visual-motor coordination, the turn-taking behaviour itself is driven by the human experimenter who reacts cooperatively towards the robot's actions.

Both examples show that a behaviour like turn-taking which is 'social' from an observer point of view need not be explicitly encoded or modelled in the control system, it can emerge from situated activity.

The internal mechanisms which I discussed in section 3.2.3 should be considered in the same way, namely as a basic capacity of socially intelligent agents, when immersed in a social situation, might give rise to social understanding. Such internal dynamics might be difficult to control, but they could provide the means for how, from the agent's point of view, the world can provide meaning. Complex forms of internal dynamics which humans are able to experience are a rich source of meaning, and can give rise to fear and happiness. But building systems with such potentials might not necessarily be desirable. Do we want our artefacts to possess internal dynamics, to express our internal states, to be scared in one situation and to be happy in another?

5.3 SIA and Human Nature

Humans adapt to technology, human cognition is shaped by behavior, 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 adapting more than the other. An example that virtual environments are tools which can change human sensorimotor coordination and self-perception (body-image) is discussed by Frank Biocca in [Bio97]. However, humans are experts in learning and adaptation, they can very flexibly get used to even very awkward interfaces (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. Thus, SIA research could learn from human factors (ergonomics) about the study of how humans and machines interact in order to design technology that work well in 'human terms' ([RH84]). Designing artificial agents which make interaction natural for humans is necessary if humans should not get used to acting (and thinking) like artificial agents. But 'humanizing' the interface and the relationships between agents and humans depends on what is meant by being 'human', and to what extent metaphors influence our conceptions of what is social and what kind of social behavior is desirable.

There is one aspect of human-style social intelligence which is in a special way 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* [WP96] violence in the context of human evolution. They suggest that the violent ‘temperament’ originates in a specific form of social organization 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 their 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 behavioral 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 ‘like-minded’ people), are these more positively, ‘peaceful’ oriented visions of human sociality appropriate or rather naive? 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 part of agent research (e.g. [PC97]). Thus, agent research (based on models from social sciences, sociology, etc.) is trying to use realistic assumptions about human social behavior. On the other hand: Humans can to a great extent chose how they want to lead their lives. Thus, the fact that at present violence still plays an important part in our society does not necessarily mean that the same will be 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). Thus, SIA agent models which draw their inspirations from human social behaviour should make explicit to which model of sociality and interaction they refer, since it cannot be taken for granted that specific forms of social behavior have a universal meaning and interpretation (see also discussion in section 7).

6 Summary

Based on the concepts presented in this paper I propose the following terminology which applies to artificial and biological agents and which can be used in the field of SIA.

- **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 of self and, via empathy, of oth-

ers. Both kinds of dynamics are two aspects emerging from the same state of being-in-the-world.

- **Autonomous agents.** Autonomous agents are entities inhabiting our world, being able to react and interact with the environment they are located in and with other agents of the same and different kind. This is a variation of Franklin and Graesser's definition ([FG97]). Biological agents, animals and plants, both their behavior 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. The same concepts can be applied to artificial agents which 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 reconstruct their individual autobiographical 'story' during their life-time by means of a dynamical memory. The autobiography reflects stories about the agents themselves as well as encounters and relationships with other agents.
- **Human social intelligence.** Humans live in individualized societies, individuals interact as 'persons', their coupling with the world consists of external (behavioral, 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 individualized, 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.
- **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). The term *artificial social intelligence* refers then to an instantiation of human-style social intelligence in artificial agents. Thus, the term social intelligence is always used in the context of human-style social interaction and behavior. 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 anonymous colony, the superorganism.

7 Conclusion

The following list of qualities of a CT approach to SIA technology might serve as general design guidelines for SIA technology. They result from the particular ‘Embodied AL’ stance which this paper has outlined. These guidelines are very general, and each single one is not new, i.e. has already played a role in designing systems in areas like software engineering (e.g. using ethnography in interactive systems design in order to assess the social context of work [HKRA95]), human-computer-interaction or computer-supported-cooperative-work. However, it is the combination of these factors which can make the big difference. Addressing these factors can make SIA systems more ‘complete’ and balanced.

1. Humans are **embodied agents**. Human understanding cannot be decoupled from the phenomenological dimension, the experiential grounding in a living body. Living through this body implies that humans are adapted to and have evolved in the real world. Experiences provided by technological means have to be grounded in the real world, but have to address the subjective, phenomenological nature of experiences. Example: Bruner ([Bru91]) mentions that time in narrative events is ‘human-time’ rather than abstract or clock time. Thus, SIA’s which are cooperating with humans e.g. as personal assistants should be able to handle both objective and subjective time in human dialogues and in the way human’s remember events and personal experiences.
2. Humans are **active agents**, they want to use their body and explore the environment. Developmental psychology points towards the crucial role of an infant’s dynamic inter-action with the environment ([HJ96]). The ontogenetical development of human cognition and intelligence is grounded in this coupling. Example: Treating users as active agents is not only relevant in interface design for children ([Sur97]), it is also relevant for normal adult users. A computer terminal poses strict constraints on bodily activity but can support ‘mental mobility’, e.g. further curiosity and creativity of users, using multi-media facilities to make work more ‘interesting’. New interfaces (like head mounted displays) are therefore not only providing new functionalities and improved efficiency of work, but meet the natural human needs of activity better than old interfaces with very few degrees of freedom for activity.
3. Humans are **individuals**, and they want to be treated as such. The viewpoints and personalities of humans always differ, no matter whether they have the same genotype or not¹⁰. Example: Approaches in which robots or software agents learn from users and mimic their behavior (e.g. see programming by demonstration/example, [Cyp93], [KDK97], [Boo98]) are resulting in products which reflect the individual behaviour or preferences of users. The degree of mimicking can range from extracting single individual features to more complex forms of imitation. Imitation is not only an efficient social learning mechanism,

¹⁰The issue of individuality and embodiment is discussed in more detail in [DC96].

but is as a mechanism by which infants get to know other persons, as objects which imitate and can be imitated, thus, objects which pass the ‘like me’ test¹¹. Recent research on imitating robots shows the relevance of imitation for social learning and communication between agents, see [HD94], [DRH⁺97], [BH97], [GMBQ97], [BD98]. Thus, imitation and social learning mechanisms might make socially intelligent agents more ‘like us’ and make them individuals.

4. Humans are **social beings**. Human intelligence can only develop in a social context, and can only be understood within a social, cultural context. The feeling of belonging, the possibility to socialize has to be provided by a CT approach of SIA. Technology can mediate between humans, it can enhance and shape their means of social interaction. Example: The effect of presence in virtual environments is enhanced by other agents inhabiting the same world, i.e. it makes the world more believable when agents are not alone (see [Sto93]). Thus, sociality is not only a prerequisite for cooperative behavior and problem-solving, and meets human needs to socialize, but can improve the acceptance of tools (like virtual environments) in the first place. Humans are, from an evolutionary perspective, primarily social beings and not rational problem solvers. However, rationality and consistency are useful concepts and result in techniques (e.g. logic, problem-solving, mental models) which have proved to be valuable in order to survive in a complex, dynamic and unpredictable environment. But the human brain and human intelligence did not evolve to let humans become good mathematicians in the first place, so information technology (as it already does) can give assistance and empower humans to be more successful in these areas. Example: Scheduling, mail-filtering and other systems which assist daily routine work can significantly reduce cognitive load.
5. Humans are **story-tellers**. Creating and reconstructing stories is at the center of human understanding. A CT approach to SIA technology can enhance the possibilities to use and re-use stories. Agents can be made more believable when put into an ‘historical’ (story) context. Example: We discussed previously that the ‘natural’ form by which humans understand, interpret and interact with the world is in terms of stories. If stories are fundamental to human (social) intelligence, then SIA’s have to be good at telling and listening to stories. ‘Stories’ are not necessarily in linguistic terms, e.g. a story can also be told in pictures or by means of non-verbal communication. In educational systems for children, presenting knowledge in terms of stories is a widely-used techniques, primarily to make the content more interesting and acceptable for children, e.g. in [Yaw97].
6. Humans are **animators**. They are biased towards ‘animating’ the world, to interpret the world in terms of intentionality and explanation. In addition to functionality a CT approach to SIA can produce believable products, so that

¹¹For discussions and pointers to research on imitation in biology, developmental psychology and robotics see [Dau95], [HG96].

they are appealing and interesting, and humans are willing to become engaged in interactions with them. Copying nature can be one way to go, e.g. building humanoid or dog-like robots ([Bro96], [FK97]). In these cases the behavior and appearance of the systems makes them spontaneously interesting. The other alternative is to design believable systems without mimicking nature ([Sen97]). The approach of mimicking nature has at least one disadvantage, namely that users associate the appearance of an agent with the intelligence or complexity of its natural model, e.g. humanoid agents which do not behave convincingly human appear unbalanced.

7. Humans are **autobiographic agents** and life-long learners. They are constantly re-telling and re-interpreting their autobiography and their interpretation of the world. They constantly learn and re-learn and can adapt to different circumstances and re-write their autobiography. Systems which assist the re-construction of autobiographical memory can therefore strengthen the development of memory, social skills and the self. Example: Jennifer Glos and Justine Cassell describe in [GC97] ‘Rosebud’ a system which uses digitally augmented ‘keepsake objects’ (here: stuffed animals) in order to enhance children’s story-telling skills, but also as memory objects of childhood. Such systems can take part in the development of an autobiography, and an individual self. And, if shared with other people, can further the culture of story-telling and the creation of social selves. According to the *social interaction hypothesis* an autobiographical memory develops gradually, children learn, by talking about memories with others, how to structure and formulate their memories in narrative form, in order to retain them in recoverable form ([Nel93]).
8. Humans are **observers**. Human perception and cognition is inherently subjective, we cannot take the stance of being an objective observer of the world. Humans are always part of the system, observations cannot be made without an observer. Example: The perspective and the historical/cultural context is important to understand human behavior and underlying motivations, goals and value systems. If an anthropologist tells us the story of a society where human sacrifice and cannibalism is central to its culture, including the practise of removal of the heart, display of the head, and wearing the skin of a flayed victim, then understanding such a behavior from a viewpoint rooted in European culture is at least difficult. Even if we get the additional information that this behavior was displayed by the Aztecs, it does not help understanding. It does help to know about the motifs and beliefs, namely that this practice “arose out of the debt owed to the gods: the consequence of not paying was no less than the end of the world” ([PG97], p. 36). A particular ‘story’ about the origin of the universe explains, and, from the Aztec point of view, fully justifies this behavior. Thus, beliefs, motifs, values which are driving the way how we perceive, understand and interact with the world, and are culturally situated. The cultural dimension of understanding others mental states, minds and behavior is stressed by Lillard ([Lil97]). She shows that theories of mind differ

across cultures and that theories based on results from European-American culture are difficult to generalize. We recognize differences when travelling to other countries, but when it comes to building technology universal aspects of social intelligence, social behavior and communication are generally assumed, e.g. generally software systems allow only minor adaptations like changing a few parameters (like adjusting language). In [OB97] O’Neill-Brown discusses the need to go beyond that step and argues for culturally adaptive agents which can adapt to cross-cultural differences in communicative behaviors.

9. A matter of **balance**. SIA design should be *balanced*, first of all in the sense of a cognitive fit between agent and human. Secondly, a balance refers to how the agent is historically grounded (autobiographic agents) so that it addresses and therefore enhances the narrative nature of human understanding. Lastly, believability of agents seems to depend on the ‘completeness’ of the design (and the design process), including social, ecological and cultural situatedness of the agent and its body.

I would like to conclude the paper by identifying application areas in which the SIA qualities mentioned above are more or less relevant. In unconstrained ‘fun’ contexts, e.g. when computer games use software agents or virtual pets, the user is him/herself already sufficiently motivated to use the tools and to become engaged in interactions. However, the user can choose one or the other product, so the most believable products with the most ‘natural’ forms of social interaction are likely to succeed. The functionality of a computer game is less constrained than e.g. the functionality of a text editor. Instead, believability of the character and the story are vital. The situation is different in business contexts, e.g. when functionality and efficiency (e.g. of an email filtering systems or a Webbrowser) are important. Thus, designing SIA technology in this area is much more constrained than for SIA game products. The degree of expressiveness and believability of SIA in this context have to fit strongly their functionality. For example, it might not be desirable to have a socially intelligent agent if it takes too much memory, processing time, or requires complex social interaction with the user until the agent can complete its task. In such a case, and I suppose for most repetitive tasks which do not require intensive user interaction, a non-social agent which is simply doing its job regardless of the user might be preferred. Thus, how can the tradeoff between efficiency and sociality be handled? SIA are desirable and preferable in contexts when their function is primarily social. Going back to our distinction in section 5.2 between biological anonymous and individualized societies, SIA should support individualized interactions, when personality, character, individual relationships are desirable. For instance, a personal assistant (a robot or a software agent) which works in close contact to a single human user, ‘cares’ about the user and which shows a human-style form of social intelligence is a desirable socially intelligent agent.

That artificial agents are becoming part of human life seems to be a ‘natural’ result of human cultural development. The future is still to come, and we can only

guess where ‘life-like’ SIA technology might lead us.

“Robbie’s chrome-steel arms (capable of bending a bar of steel two inches in diameter into a pretzel) wound about the little girl gently and lovingly, and his eyes glowed a deep, deep red.”[Asi68], p. 36

We and the societies we are living in are changing, and so do our conceptions of being ‘social’. Thus, the process and products of socially intelligent agent design might be seen as a mirror which reflects our own sociality.

References

- [Aro94] Elliot Aronson. *The social animal*. W.H. Freeman and Company, New York, 1994.
- [Asi68] Isaac Asimov. *I, Robot*. Grafton Books, Collins Publishing Group, London, 1968.
- [Bar32] F. C. Bartlett. *Remembering – A study in experimental and social psychology*. Cambridge University Press, 1932.
- [Bat94] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
- [BD98] Aude Billard and Kerstin Dautenhahn. Grounding communication in autonomous robots: an experimental study. to appear in: *Robotics and Autonomous Systems, Special Issue on “Quantitative Assessment of Mobile Robot Behaviour”*, guest editors: U. Nehmzow, M. Recce, D. Bisset, 1998.
- [BH97] Aude Billard and Gillian Hayes. Learning to communicate through imitation in autonomous robots. In *Proceedings of ICANN97, 7th International Conference on Artificial Neural Networks*, pages 763–768. Springer-Verlag, 1997.
- [Bio97] Frank Biocca. The cyborg’s dilemma: Embodiment in virtual environments. In Jonathon P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska, editors, *Proceedings of the Second International Conference on Cognitive Technology*, pages 12–26. IEEE Computer Society Press, 1997.
- [Boo98] Gary Boone. Concept features in re:agent, an intelligent email agent. In *to appear in: Second International Conference on Autonomous Agents (Agents ’98), Minneapolis/St. Paul, May 9-13, 1998*, 1998.
- [Bro91a] Rodney A. Brooks. Intelligence without reason. In *Proc. of the 1991 International Joint Conference on Artificial Intelligence*, pages 569–595, 1991.

- [Bro91b] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.
- [Bro96] Rodney Brooks. Behavior-based humanoid robotics. In *Proc. 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 96*, pages 1–8, 1996.
- [Bro97] Rodney A. Brooks. personal communication. 2nd International Conference on Cognitive Technology (CT'97), Aizu, Japan, August 28, 1997, 1997.
- [Bru91] Jerome Bruner. The Narrative Construction of Reality. *Critical Inquiry*, 18(1):1–21, 1991.
- [BW88] R. W. Byrne and A. Whiten. *Machiavellian intelligence*. Clarendon Press, 1988.
- [Byr95] R. Byrne. *The thinking ape, evolutionary origins of intelligence*. Oxford University Press, 1995.
- [Con96a] Martin A. Conway. Autobiographical knowledge and autobiographical memories. In David C. Rubin, editor, *Remembering our past. Studies in autobiographical memory*, pages 67–93. Cambridge University Press, 1996.
- [Con96b] Martin A. Conway. Shifting sands. *Nature*, 380:214, 1996.
- [CS92] D. L. Cheney and R. M. Seyfarth. Précis of how monkeys see the world. *Behavioral and Brain Sciences*, 15:135–182, 1992.
- [CvdV97] Dolores Canamero and Walter van de Velde. Socially emotional: Using emotions to ground social interaction. In *Socially Intelligent Agents*, pages 16–21. AAAI Press, Technical report FS-97-02, 1997.
- [Cyp93] A. Cypher, editor. *Watch what I do: Programming by demonstration*. MIT Press, 1993.
- [Dau95] Kerstin Dautenhahn. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16:333–356, 1995.
- [Dau96] Kerstin Dautenhahn. Embodiment in animals and artifacts. Working Notes AAAI 96 Symposium on Embodied Action and Cognition, 1996.
- [Dau97] Kerstin Dautenhahn. I could be you – the phenomenological dimension of social understanding. *Cybernetics and Systems*, 25(8):417–453, 1997.
- [Dav96] Glorianna Davenport. Smarter tools for storytelling. are they just around the corner? *IEEE Multimedia*, 4(1):10–14, 1996.

- [Dav97] Glorianna Davenport. Encounters in dreamworld: a work in progress. In Roy Ascott, editor, *Proc. of the First International CAiiA Research Conference*, 1997.
- [DC96] Kerstin Dautenhahn and Thomas Christaller. Remembering, rehearsal and empathy - towards a social and embodied cognitive psychology for artifacts. In Sean O’Nuallain and Paul Mc Kevitt, editors, *Two sciences of the mind. Readings in cognitive science and consciousness*, pages 257–282. John Benjamins North America Inc., 1996.
- [Den87] Daniel C. Dennett. *The intentional stance*. MIT Press, 1987.
- [DRH⁺97] J. Demiris, S. Rougeaux, G. M. Hayes, L. Berthouze, and Y. Kuniyoshi. Deferred imitation of human head movements by an active stereo vision head. In *Proceedings of the 6th IEEE International Workshop on Robot Human Communication, Sendai, Japan, Sept. 1997*, pages 88–93. IEEE Press, 1997.
- [Dun93] R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–735, 1993.
- [FG97] 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.
- [FK97] Masahiro Fujita and Koji Kageyama. An open architecture for robot entertainment. In W. Lewis Johnson, editor, *Proc. of the First International Conference on Autonomous Agents, Marina del Rey, CA, USA, February 5-8*, pages 435–442, 1997.
- [FSR97] Adam Frank, Andrew Stern, and Ben Resner. Socially intelligent virtual petz. In *Socially Intelligent Agents*, pages 43–45. AAAI Press, Technical report FS-97-02, 1997.
- [GC97] Jennifer W. Glos and Justine Cassell. Rosebud: A place for interaction between memory, story, and self. In Jonathon P. Marsh, Christopher L. Nehaniv, and Barbara Gorayska, editors, *Proceedings of the Second International Conference on Cognitive Technology*, pages 88–97. IEEE Computer Society Press, 1997.
- [GCM97] Stephen Grand, Dave Cliff, and Anil Malhotra. Creatures: Artificial life autonomous software agents for home entertainment. In *Proc. First International Conference on Autonomous Agents (Agents ’97), held in Marina del Rey in February 1997*. ACM, 1997.
- [Gee88] Karl D. S. Geeste. *Stümpke’s Rhinogradentia. Versuch einer Analyse*. Gustav Fischer Verlag, Stuttgart, 1988.

- [GMBQ97] P. Gaussier, S. Moga, J. P. Banquet, and M. Quoy. From perception-action loops to imitation processes: A bottom-up approach of learning by imitation. In *Socially Intelligent Agents*, pages 49–54. AAAI Press, Technical report FS-97-02, 1997.
- [GMM97] Barbara Gorayska, Jonathon Marsh, and Jacob L. Mey. Putting the horse before the chart: formulating and exploring methods for studying Cognitive Technology. In Jonathon P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska, editors, *Proceedings of the Second International Conference on Cognitive Technology*, pages 2–9. IEEE Computer Society Press, 1997.
- [HC80] A. S. Hornby and A. P. Cowie. Oxford advanced learner’s dictionary of current english. Cornelson and Oxford University Press, 11th Edition, 1980.
- [HD94] Gillian Hayes and John Demiris. A robot controller using learning by imitation. In *Proc. International Symposium on Intelligent Robotic Systems, Grenoble*, pages 198–204, 1994.
- [HG96] Cecilia M. Heyes and Bennett G. Galef. *Social Learning in Animals: The Roots of Culture*. Academic Press, 1996.
- [HJ96] Horst Hendriks-Jansen. *Catching ourselves in the act: situated activity, interactive emergence, evolution, and human thought*. MIT Press, Cambridge, Mass., 1996.
- [HKRA95] John Hughes, Val King, Tom Rodden, and Hans Anderson. The role of ethnography in interactive systems design. *Interactions*, 4:57–65, 1995.
- [HRvG97] Barbara Hayes-Roth and Robert van Gent. Story-Making with Improvisational Puppets. In W. Lewis Johnson, editor, *Proc. of the First International Conference on Autonomous Agents*, pages 1–7, 1997.
- [KDK97] Volker Klingspor, John Demiris, and Michael Kaiser. Human-robot-communication and machine learning. *Applied Artificial Intelligence Journal*, 11:719–746, 1997.
- [KGR96] David Kinny, Michael Georgeff, and Anand Rao. A methodology and modelling technique for systems of BDI agents. In Walter Van de Velde and John W. Perram, editors, *Agents Breaking Away, Proc. of 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW’96, Eindhoven, The Netherlands, January 1996*, pages 56–71, 1996.
- [KPI96] K. Kawamura, R. T. Pack, and M. Iskarous. Design philosophy for service robots. *Robotics and Autonomous Systems*, 18:109–116, 1996.

- [Kus97] Nicholas Kushmerick. Software agents and their bodies. *Minds and Machines*, 7(2):227–247, 1997.
- [Lan89] Christopher G. Langton. Artificial 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 1–47, 1989.
- [LB97] A. Bryan Loyall and Joseph Bates. Personality-rich believable agents that use language. In W. Lewis Johnson, editor, *Proc. of the First International Conference on Autonomous Agents*, pages 106–113, 1997.
- [Ld95] M. Luck and M. d’Inverno. A formal framework for agency and autonomy. In *Proceedings of the First International Conference on Multi-Agent Systems*, pages 254–260. AAAI Press/MIT Press, 1995.
- [Lev92] Steven Levy. *Artificial Life, The quest for a new creation*. Penguin Books, 1992.
- [Lil97] Angeline S. Lillard. Other folks’ theories of mind and behavior. *Psychological Science*, 8(4):268–274, 1997.
- [Mae90] Pattie Maes, editor. *Designing autonomous agents: theory and practice from biology to engineering and back*. The MIT Press, 1990.
- [Mat95] Maja J. Mataric. Issues and approaches in design of collective autonomous agents. *Robotics and Autonomous Systems*, 16:321–331, 1995.
- [Mat97] Michael Mateas. Computational subjectivity in virtual world avatars. In *Socially Intelligent Agents*, pages 43–45. AAAI Press, Technical report FS-97-02, 1997.
- [MNG97] Jonathon P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska. Cognitive technology, humanizing the information age. In *Proceedings of the Second International Conference on Cognitive Technology*, pages vii–ix. IEEE Computer Society Press, 1997.
- [Nel93] Katherine Nelson. The psychological and social origins of autobiographical memory. *Psychological Science*, 4(1):7–14, 1993.
- [OB97] Patricia O’Neill-Brown. Setting the stage for a culturally adaptive agent. In *Socially Intelligent Agents*, pages 93–97. AAAI Press, Technical report FS-97-02, 1997.
- [PC97] Michael Prietula and Kathleen Carley. Agents, trust and organizational behavior. In *Socially Intelligent Agents*, pages 146–149. AAAI Press, Technical report FS-97-02, 1997.

- [Pen95] Simon Penny. The pursuit of the living machine. *Scientific American*, 9:172, 1995.
- [Pfe95] Rolf Pfeifer. Cognition – perspectives from autonomous agents. *Robotics and Autonomous Systems*, 15:25–46, 1995.
- [PG97] Holly Peters-Golden. *Culture Sketches. Case Studies in Anthropology*. The McGraw-Hill Companies, Inc., 1997.
- [PP95] 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.
- [Pre97] Erich Prem. Epistemological aspects of embodied artificial intelligence. *Cybernetics and Systems*, 28(5):iii–ix, 1997.
- [Ray92] Thomas S. Ray. An approach to the synthesis of life. In C. G. Langton, C. Taylor, and J. D. Framer, editors, *Proc. of the Second Artificial Life Workshop*, pages 371–408, 1992.
- [Ray94] Thomas S. Ray. An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life*, 1:179–209, 1994.
- [Rei97] W. Scott Neil Reilly. A methodology for building believable social agents. In W. Lewis Johnson, editor, *Proc. of the First International Conference on Autonomous Agents*, pages 114–121, 1997.
- [RH84] Richard Rubinstein and Harry M. Hersh. *The Human Factor. Designing Computer Systems for People*. Digital Press, 1984.
- [RM95] 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, *Understanding other minds, perspectives from autism*, chapter 7, pages 139–152. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995.
- [Rot94] Gerhard Roth. *Das Gehirn und seine Wirklichkeit*. Suhrkamp, 1994.
- [RS97] Charles Rich and Candace L. Sidner. COLLAGEN: When Agents Collaborate with People. In *Proceedings of the First International Conference on Autonomous Agents*, pages 284–291, 1997.
- [SA77] R. C. Schank and R. P. Abelson. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Erlbaum, Hillsdale, NJ, 1977.
- [SB95] Luc Steels and Rodney Brooks, editors. *The artificial life route to artificial intelligence: building embodied, situated agents*. Lawrence Erlbaum Assoc., 1995.

- [Sen97] Phoebe Sengers. Towards socially intelligent agent building. In *Socially Intelligent Agents*, pages 125–130. AAAI Press, Technical report FS-97-02, 1997.
- [Sha97] Mike Sharples. Storytelling by Computer. *Digital Creativity*, 8(1):20–29, 1997.
- [Sim63] G. G. Simpson. Review. *Science*, 140(3567):624–625, 1963.
- [SJA91] 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.
- [Ski74] Burrhus F. Skinner. *About behaviorism*. Random House, New York, 1974.
- [St9] Harald Stümpke. *Bau und Leben der Rhinogradentia*. Gustav Fischer Verlag, Stuttgart, 1989.
- [Sto93] V. E. Stone. Social interaction and social development in virtual environments. *Presence*, 2(2):153–161, 1993.
- [Sur97] Jane Fulton Suri. Challenges and opportunities in designing for preschool-aged children. *Interactions, Special issue on User Interfaces for Young and Old*, pages 37–39, 1997.
- [Tat97] A. Tate. Mixed-Initiative Interaction in O-Plan. In *Working notes of the AAAI Spring Symposium on Computational Models of Mixed-Initiative Interaction*, 1997.
- [TJ81] F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, 1981.
- [TP97] Robert Trappl and Paolo Petta, editors. *Creating personalities for synthetic actors*. Springer, 1997.
- [vdV97] Walter van de Velde. Co-Habited Mixed Realities. In *Proceedings of the IJCAI'97 Workshop on Social Interaction and Communityware, Nagoya, Japan, August 1997*, 1997.
- [Wat24] John B. Watson. *Behaviorism*. University of Chicago Press, Chicago, 1924.
- [Wat95] Stuart Watt. The naive psychology manifesto. The Open University, Knowledge Media Institute, Technical Report, number KMI-TR-12, 1995.
- [Weg97] Peter Wegner. The paradigm shift from algorithms to interaction. *CACM*, 40(5):80–91, 1997.

- [WJ95] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.
- [Wor96] Robert P. Worden. Primate social intelligence. *Cognitive Science*, 20(4):579–616, 1996.
- [WP96] Richard Wrangham and Dale Peterson. *Demonic males: apes and the origins of human violence*. Houghton, 1996.
- [WPAK97] D. M. Wilkes, R. T. Pack, A. Alford, and K. Kawamura. HuDL, A Design Philosophy for Socially Intelligent Service Robots. In *Socially Intelligent Agents*, pages 140–145. AAAI Press, Technical report FS-97-02, 1997.
- [WR97] Philip F. Webb and Jeremy I. Robson. A human factors based approach to the development of advanced teleoperated robotic systems. In C. A. Czarnecki, editor, *Proc. of Workshop on Recent Advances in Mobile Robots, July 1st 1997, De Montfort University, Leicester*, pages 67–71, 1997.
- [Wye95] Robert S. Wyer. *Knowledge and memory: the real story*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995.
- [Yaw97] Mitchell A. Yawitz. Building a time and a space. *Interactions, Special issue on User Interfaces for Young and Old*, pages 37–39, 1997.