Getting to Know Each Other – Artificial Social Intelligence for Autonomous Robots

Kerstin Dautenhahn
VUB AI-Lab, Belgium
kerstin@arti.vub.ac.be

the final copy of this paper appeared as: K. Dautenhahn (1995), Getting to know each other – artificial social intelligence for autonomous robots, Robotics and Autonomous Systems 16, pp 333-356

Abstract

This paper proposes a research direction to study the development of 'artificial social intelligence' of autonomous robots which should result in 'individualized robot societies'. The approach is highly inspired by the 'social intelligence hypothesis', derived from the investigation of primate societies, suggesting that primate intelligence originally evolved to solve social problems and was only later extended to problems outside the social domain. We suggest that it might be a general principle in the evolution of intelligence, applicable to both natural and artificial systems. Arguments are presented why the investigation of social intelligence for artifacts is not only an interesting research issue for the study of biological principles, but may be a necessary prerequisite for those scenarios in which autonomous robots are integrated into human societies, interacting and communicating both with humans and with each other. As a starting point to study experimentally the development of robots' 'social relationships', the investigation of collection and use of body images by means of imitation is proposed. A specific experimental setup which we use to test the theoretical considerations is described. The paper outlines in what kind of applications and for what kind of robot group structures social intelligence might be advantageous.

1 Introduction

Social skills have been mostly regarded as a 'side-effect' or even neglected so far in those research areas which are dealing with the construction of intelligent artifacts, namely artificial intelligence, robotics and also in the relatively new research field artificial life (alife). Surprisingly, for a long time in natural sciences (e.g. primatology, child development, psychology), broad discussions have been taking place about the role of 'the social factor' in the development of intelligence. This paper's goal is to suggest that these ideas should also find their way into the 'sciences of the artificial'.

Most research in robotics aims at developing robots with a domain-specific 'technical intelligence', e.g. building robots which can use effectively and with high precision a manipulator or identify specific patterns. In the same way so-called 'robot contests', which are held at different laboratories world-wide, are dominated by events in which single robots have to cope with technical tasks, e.g. navigation or object manipulation. The main focus of interest is *competition* between robots, in its extreme case in a wrestling-like situation. This might be due to the 'play-like' origin of these activities, but there might be a certain feedback to scientific research which sometimes seems to be at least *inspired* by these events.

In the same way, artificial intelligence research has focussed on isolated non-social aspects of intelligence (e.g. logic, automated theorem proving or diagnostic reasoning). In principle, 'social intelligence' need not necessarily be interesting in robotics when other technical solutions are possible. But, as we will show, the scenarios being imagined for 'real world' robots (e.g. welfare robots, domestic robots, robots acting in teamwork to solve a common problem) require to a high extent aspects of communication and cooperation between robots and between robots and humans. Especially in those situations where the robots should support humans in jobs with many social contacts (e.g. working as a filling station attendant), then in addition to a domain-specific 'laser beam intelligence' (like the capability to perform fast calculations), the robots should be able to effectively (i.e. human-like) communicate with humans. Such a competence might be much more important for the acceptance and 'social integration' of the robots into human societies than a robot's outer appearance. Moreover, if the acquisition of social intelligence will be proved to be one important prerequisite for the development of non-social kinds of intelligence (see section 3), then artifacts (e.g. robots) *must* have social skills, even if their application domain is dominated by non-social activities. Derived from the impression that many of today's robots have a highly complex 'body'[1] (referring to mechanics, electronics, sensors and so on) but still a very primitive kind of intelligence, the acquisition of 'social intelligence' might be an interesting path towards filling this gap.

Section 2 describes scenarios for autonomous robots in real world applications which point to the need to develop 'social skills' for interaction and communication among robots and between robots and humans, i.e. the need to develop some kind of 'social intelligence'. Section 3 discusses experimental findings from primatology studying the development of intelligence in primate societies. Based on this data from the study of natural societies, we draw conclusions and formulate hypotheses about the social life of present-day robots. Section 4 outlines the direction from

[1]We should note that every time we use terms which are originally defined in natural sciences and refer to biological systems, these terms are always used in a metaphorical sense. We are aware of the problem that the reader might confuse some of these terms with their everyday meaning which might be quite different and misleading. But since this is a strongly interdisciplinary paper, it is not possible to define each technical term, so we take the risk of misinterpretation.

natural to artificial social intelligence, sketching two research fields dealing with closely related issues. Sections 4.1 and 4.2 describe two mechanisms which we propose to study in order to approach social intelligence for artifacts, namely imitation and body images. This paper is not a technical paper, but, to go beyond a mere theoretical discussion, we describe in section 5 our experimental setup designed to study 'artificial social intelligence'. Section 6 discusses situations in which social intelligence may be useful for autonomous robots. In addition, we outline different kinds of relationships between robot and human societies. The last section (7) sketches our conception of a future development of autonomous robots in relation to the approach presented in this paper.

2 Scenarios for autonomous robots

Recently, different research proposals and studies have been initiated outlining scenarios for applications of autonomous robots, mainly focussing on service robots, i.e. robots belonging to our daily life, and therefore stressing the need for robots to have interaction and communication abilities (e.g. [Ots93],[Ge94],[FhG94]). Additionally, concrete research activities are under way, aiming at building 'humanoid robots' which should interact with humans (e.g. [BS93],[IS94]). While there exist efficient methods for the control of robots in well-structured (often man-made and static) environments (e.g. automation processes), there seems to be a certain kind of 'gap' concerning robots which, due to the complexity of the task and/or the environment, should show a certain degree of 'intelligence' with respect to flexibility, adaptivity, robustness and 'autonomy', the latter especially important for mobile robots. But the service robots which are supposed to become members of our daily life do not only face the problem of coping with the 'inanimate' environment, but also with the 'animate environment'.

We now want to make clear for more technical reasons (functionality and user acceptance) why we think that dealing effectively with the 'animate'[2] environment requires a sort of 'social intelligence'. The main idea of using 'personal robots' in service applications (e.g. as household assistants or welfare robots) is not new at all, e.g. it has already been described by Ichiro Kato in 1991 ([Kat91][3]). Kato gave a specification of future humanoid welfare robots including one point which is also important for our argumentation, namely "'They are capable of adapting to human motion and feelings." In our view this points to the *individual* character of the robots' behavior. For instance a robot helping a human with a specific handicap might benefit from having some 'generalized' knowledge about both human movements and the specific disease the individual is suffering from, but the robot could

[2]We use this term most often referring to 'other autonomous systems', including other robots, too.

[3]Kato himself mentioned that the idea has been expressed much earlier by M. W. Thring in 1964.

not be helpful unless he could adapt to the 'individual' bodily and mental characteristics of the human. If robots have long-term contact with a person, it might be desirable, in addition to robots providing technical support, to have them develop a 'social skill' and individual relationships. This would enhance the performance of the robots (e.g. anticipating people's behavior) and would be desirable for most humans who prefer being treated as an 'individual' rather than an 'anonymous' patient. For the development of *individual relationships* to humans, it does not suffice to store information about outer appearance, behavior or goals and beliefs of the interaction partner, since 'social skills' cannot be defined as the 'rational manipulation' of others like chess pieces, but are strongly related to individual feelings, emotional involvement, and empathy[4].

In the context of 'nursery robots', the problem of safety concerning robot actions is intuitively clear. Safety might be enhanced if robots would somehow 'care about' the welfare of others, particularly the patient's welfare. But also in other application fields of autonomous robots: not each robot can continuously be checked for whether its behavior is 'safe'. Instead, it is desirable that the robot itself controls its 'safety' autonomously. So it has to know, at least to a certain extent, about the consequences of its behavior to others, how others judge the robot's behavior, how they want it to behave, and so on, all points being part of the 'social management' practice.

Therefore, we suggest that robots which search their way 'out of factories and into human everyday life' which have close contacts to humans who are (1) highly individual with respect to bodily features and behavior and (2) are highly social beings, must develop social skills and a kind of 'social intelligence'. This is necessary for (1) fulfilling their task, (2) being accepted by humans and (3) being integrated into human society.

Sections 6 and 7 at the end of this paper will go into more detail concerning robot-robot and human-robot relationships. We now want to explain why social intelligence for robots might be interesting even if the robots are used in applications where they are nearly all the time dealing with non-social tasks. Hints are coming from studies in natural sciences.

[4]Throughout this paper we often use mental or cognitive terms in the context of artifacts. We are aware of the important distinction between the cognitive domains of the human observer and the cognitive domain of other natural or artificial systems. Stuart Watt ([Wat95]) impressively pointed out the relationships between our human, anthropocentric point of view and our ability to recognize intelligence in other objects. Anthropomorphism is not just a possible 'danger' which is involved in our relationship towards artifacts, but (possibly) we are not able to avoid it and 'step outside our humanity'. Therefore Watt proposes to study how people ascribe mental states to other people, to animals, and to machines. Although many people in alife research (including ourselves) discuss about things which should be 'really important' to the life of autonomous agents, our scientific understanding about this is deeply shaped by our life as (social) human beings. And this is why the argumentation presented in this paper can only be understood by human readers.

3 Natural social intelligence

The following citation (found in [Aro94]) shows that the idea to emphasize the social aspects of human nature is not an invention of scientists in the twentieth century.

> Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something in nature that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god. (Aristotle, *Politics*, c. 328 B.C.)

During the last decades, the social factor has come under close study in scientific discussions about the origin of human intelligence. The *social intelligence hypothesis*, which is also called the *Machiavellian intelligence hypothesis*, goes back to the ideas of several researchers, namely Chance, Humphrey, Jolly, Kummer and Mead ([CM53], [Jol66], [Hum76], [KG85]). According to the social intelligence hypothesis, (see [BW88] or [Byr95], chapter13 for an overview) "'...although most research on animal and human intellect has focussed on how intelligence deals with the physical or technical world (and the very concept of intelligence has been shaped by this), in reality intelligence is applied also in dealing with other individuals.", whereby primates use the conspecific as a 'social tool'([WB88]). This hypothesis can also be put in stricter terms, namely primate intelligence "'originally evolved to solve social problems and was only later extended to problems outside the social domain" ([CS92]). The idea that the "'...complexity of *social* life was the prime selective agent of primate intelligence" ([KG85]) is sketched in fig. 1.

To follow these ideas, in contrast to non-human primates who are able to handle complex social problems in a kind of 'laser beam' (domain specific) intelligence, humans are able to transfer and adapt knowledge from one domain to the other. Cheney and Seyfarth ([CS92]) state that "'Monkeys...do not know what they know and cannot apply their knowledge to problems in other domains." They suggest non-human primates as good primatologists, while humans can become good natural scientists. An impressive argumentation about the origin and nature of the 'Homo psychologicus' is given by Nicholas Humphrey ([Hum84]) which motivates the reader to think about the possible social origins of his or her daily life behavior, even in those areas which are supposed to be basically something 'different' (e.g. scientific experimentation). In its extreme case, this would mean that even those human behaviors which are attributed to 'rationality' have a social basis (e.g. mathematical thinking). In a similar way, Daniel Bullock ([Bul83]) suggests that the highly elaborated human ability of symbolization, which seems to be an important characteristic of human cognition, stems from social interactions. "'Symbolization is not the isolated thinker's act of representing a symbol to him- or herself." Instead,

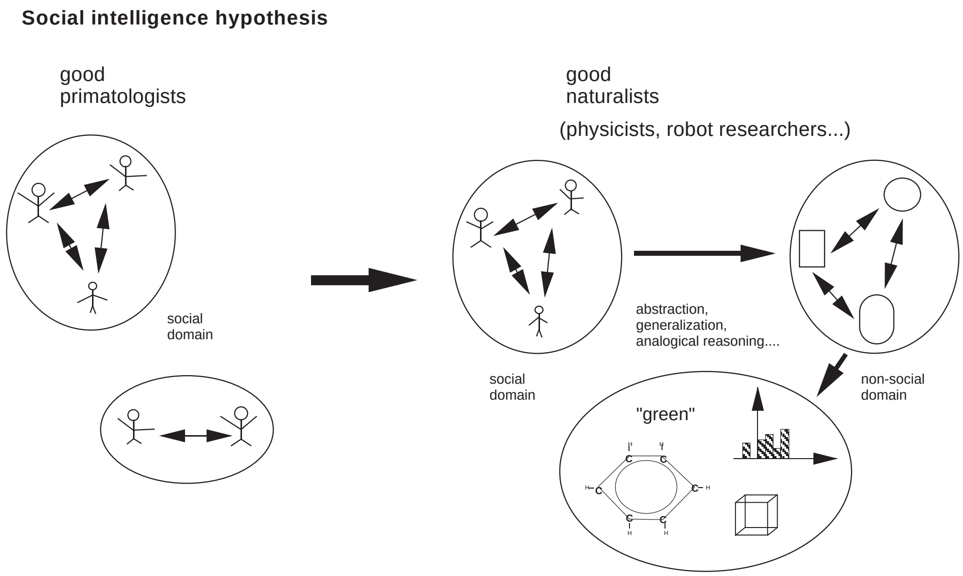

Figure 1: The 'social intelligence hypothesis': from social intelligence to abstract problem solving. See text for further explanation.

symbolization is a social act of agreeing, "'of establishing a social convention - something that logically requires a minimum of two agents." In a following paper, Bullock ([Bul87]) suggests that human intelligence as such, since it depends strongly on social embeddedness, should be thought of as a "'socially-distributed phenomenon". Bullock [Bul87] presents a 'social theory of intellectual development', stressing the *modes of interaction* which are available to a species or group member as the most important constituents of intelligence. He presents a 'convergence rate hierarchy' in which, at each of the eleven levels, the species repertoire is enlarged by a new factor, allowing faster convergence toward new adaptations (i.e. enhancing 'intelligence'). The hierarchy starts with natural selection and is step by step extended, e.g. by reflex conditioning (level 2), exploratory play (level 5), constructive imitation (level 6), or writing (level 11). According to Bullock, the addition of 'constructive imitation' to this repertoire of interactions as the most dominant mode is a crucial step in intellectual development, because it can be seen as the threshold beyond which intelligence becomes a socially distributed phenomenon, i.e. "'...the individual's adaptive power is best measured by looking jointly at the individual's imitative propensity and at the groups within which the individual interacts and freely shares information." These ideas are closely related to the approach presented in this paper, i.e. stressing the 'social roots of intelligence', and also by pointing out *imitation* as a crucial factor in the intellectual development. It is exactly this factor, namely imitation, which we choose to investigate for robot interactions (section 4.1 gives further reasons for this).

Complex social interactions evolve over time and the social relationships of an individual at a given point in time are the results of all social interactions in which the agent has been involved up to this point. Thus the ontogeny of an agent is not only important because the learning algorithms for the acquisition of certain skills take time to 'mature'. Also mental experiences are collected during life-time and contribute to the development of an individual character. If some experiences were selectively deleted, a behavior which might be called 'inconsistant' could emerge. Natural agents are embedded in space, but also in time. According to [Ros93], the aspect of time and its fundamental relationship to memory and consciousness have been disregarded too long in the classical neurologists' models for the brain. Moreover, the individual mental ontogeny and the development of social contacts are strongly interrelated with the development of the physical body of the agent. The idea of 'embodiment of social behavior' is clarified in section 4.2.

4 Approaching artificial social intelligence

Many approaches in artificial life to groups of physical robots, which take into consideration interactions between robots, prefer the simulation of social insect societies which are anonymously organized societies without individual relationships ([TGGD91], [DGF+91], [KZ94]). The individuals interact only for cooperation, tackling a problem which cannot be solved by a single agent. Other approaches take other agents only into consideration as moving obstacles or as competitors for limited resources. Non-social interactions also dominate approaches in distributed artificial intelligence and multi-agent systems on collective agents (see [Mat95] for an overview).

Based on results from the study of natural societies and especially influenced by the social intelligence hypothesis, we formulate the following hypotheses about the social life of present-day robots.

- **Robots as *Kaspar Hauser* animals.** We would state, that most present-day robots behave like animals which passed social deprivation experiments and grew up in isolation. But if, as described above, domain-independent intelligence evolved out of social-living species, why do we think that artificial agents, reared in isolation, will ever evolve a glimpse of human-like 'intelligence' and will be able to communicate effectively with members of the same or different species? Instead, robots should, comparable to the normal development of a child, 'grow up' in a social context. Research aiming at the development of intelligent artifacts should not only focus on solving problems with the dynamics of the inanimate environment, but it should also take into account social dynamics.

- **Robots without an ontogeny**. If artificial agents should resemble living beings, they should not have a reset-button: One should not only think of phylogeny when trying to simulate natural evolution of a species or the survival of an individual. Artificial agents will never show an elaborated 'mental life' if they do not have the chance to have an individual ontogeny like all natural agents. Only an agent without a reset-button has the chance to become an 'individual'. The fact that all individuals develop on a 'species specific' time-scale is intuitively known by everybody, but has had less impact on research on robots[5].

- **Disembodied robots**. We would state that it is not possible to fulfill the embodiment criterion for robots (see [Bro91]), i.e. that no robot will ever evolve a concept of a body or personal self (both for itself and for others), if we do not include the ontogeny of body concepts (see section 4.2), i.e. the collection of body images through social interactions. We suppose that it is not enough to use robots with a 'real body', unless the robots themselves actively *use* their bodies and develop a kind of 'body conception'[6].

What implications do result from the study of natural societies for robotics or alife research? Maybe the most interesting point is a corroboration of the social intelligence hypothesis. This would be the case if the process how domain-independent intelligence can emerge in a social context could be modeled. We hypothesize, that the social intelligence hypothesis is also relevant for intelligent life not based on the biological substrate, namely the cell. It might be a general principle for the evolution of intelligence. We do not state that the social factor is the only one relevant at this point. Most probably the evolution towards primate intelligence was influenced by several interdependent factors, including ecological factors like food ([BC92], [Mil93], [Fis93]).

In more and more technical domains which seemed to be typical and ideal application fields for automation, e.g. tasks with include a high degree of repetitive tasks like graphics editing or robot assembly tasks, there seems to have occurred a change in paradigm from 'full automation' to 'human-centered' problem solving, using humans explicitly as 'teachers' for machines (see 'programming by demonstration' approaches, [Cyp93]), resulting in a kind of man-machine symbiosis (see [WMMH92]).

[5] The 'lifelong learning' approach ([TM93]) explicitly faces this problem in the case of learning robot control. Additionally, there are already first steps towards evolving robots' hardware, see recent *embryonics* ([MPM+94]) or *evolvable hardware* approaches to producing hardware in a biological-like manner.

[6] In [DC95] we proposed to use a dynamic systems approach towards the development of a theory of embodied artificial cognition. In our view cognition cannot be studied isolated from the body, because cognitive abilities result from the morphogenesis of the body and its interactions with the social and non-social environment (see structural coupling, described e.g. in [MV87]). The body is not a fixed and pregiven 'actuator device', but it is a dynamic and ontogenetically evolving entity.

This tendency of an increasing cooperation between humans and artifacts is especially pushed forward by work on 'intelligent software agents' in human-computer interaction, since, according to Doug Riecken "'The basic idea of agent research is to develop software systems which *engage and help* all types of end users." ([Rie94]). This results e.g. in research on 'socially' rooted aspects like believability, emotion, or collaboration of agents (e.g. [Bat94]).

As an appropriate way to start with the investigation of artificial social intelligence, we focus on the development of individual interactions between robots. We decided to concentrate on two mechanisms which are discussed to be crucial for the development of individual interactions and social relationships: imitation and the collection of body 'images'. The following two sections describe these phenomena and stress those aspects which are relevant for our conception of 'social robots'.

4.1 Imitation

A common definition of imitation goes back to Thorpe's definiton of 'true imitation' which is seen as the "'copying of a novel or otherwise improbable act or utterance, or some act for which there is clearly no instinctive tendency" ([Tho63]). According to [Moo92], "'the capacity to imitate the movements of others...is among the least common and most complex forms of learning". In this way, imitation can be classified as a social learning method. It is necessary to clearly separate imitation from other 'imitation-like' phenomena which can be found in social living groups, e.g. contagion or normal maturation. According to the critical analysis of [Moo92], there is experimentally supported evidence for imitation only in certain primates (humans and chimpanzees), cetaceans (whales and dolphins) and parrots. All these species live in societies with individual relationships. The recognition of other individuals ('conspecifics') is necessary as a means to control the interactions, to predict the behavior of the 'conspecific', and to develop complex social relationships like 'aversion' or 'attachment'.

Since imitation requires the concentration upon one individual, we suppose that it evolved in animal societies where individual contacts between conspecifics occur. [MM92] hypothesized that young infants use imitation to "'communicate with 'persons' as opposed to 'things', and to "'probe the identity of a person", i.e. imitating movements are used to clarify ambiguities about the identity of persons. The judgement 'Here is something like me' should enable the individual to distinguish between animate and inanimate objects and therefore is necessary to build up interpersonal relations. In this way, for young infants, persons are defined as entities which pass the 'like me' test, i.e. , "'entities that can be imitated and also who imitate me". According to Meltzoff and Gopnik ([MG93]), imitation is an important means of learning socially relevant movements (especially facial movements) in primates. Therefore imitation is supposed to be an important mechanism in child development which is necessary for the development of individual contacts and social relationships.

Surprisingly, up to the present there is no commonly accepted definition of imitation available. Mitchell ([Mit87]) recently gave a formal definition of imitation and provided an comparative-developmental framework with five sequential, but hierarchically related levels of imitation, referring to five types of processes that result in imitation. Since this framework is applicable among species and machines and could also contribute to an investigation of imitation with artifacts, the next paragraph describes the framework in more detail.

Mitchell defines imitation as follows "'...imitation occurs when something C (the copy) is produced by an organism and/or machine, where: C is similar to something else M (the model); registration of M is necessary for the production of C; and C is designed to be similar to M.". First-level imitation is mimicry, "'achieved through morphogenesis and evolutionary processes of selection", including behavioral mimicry. The imitator and the model need not experience each other. At level two, M can influence C. The behavior of the imitator is controlled by an 'open program', it recreates the outer appearance or the actions of the model following perception. 'Open' means that the program only works if the imitator has interacted with the model. Since there is no perceptual tracking, the imitative behavior can be deferred. At third-level imitation, the program producing the imitation behavior can be modified in relation to the model, i.e. the imitator dynamically strives to achieve correspondence with the model. Fourth-level imitation means that the imitator controls the relationship between M and C. In order to achieve greater similarity, the imitator modifies and varies certain aspects of C. "'The imitator becomes the programmer of its own behavior, and recognizes the copy as a copy." At this level, the imitators are self-consciously imitative; they "'recognize that they are imitating", they need self-awareness. This should be done by humans, chimpanzees and dolphins. At the fifth level, the programming is planned by the imitator who "'changes its imitation in relation to its knowledge of another organism's perception". The imitator is not only self-consciously, but also other-consciously, imitating. Mature humans are mentioned as experts at fifth-level imitation, resulting e.g. in deception and art, but dolphins and chimpanzees might also be capable of fifth-level imitation. The development from level one to level five is not based on the development of internal representations of the model, but each level characterizes a particular type of 'awareness' necessary for imitation: no awareness of the imitation act (level 1, mimicry), awareness of the model, awareness of differences between C and M, awareness of the imitating act, or awareness of another's awareness (level 5).

Imitation is closely related to a behavior which is important for the development of 'intelligent' behavior and which has been usually ignored so far in research on artificial agents: play. Play is only found in highly evolved vertebrates, such as mammals and birds. Usually, it is restricted to young animals, but in some species (e.g. primates and whales) it can also be found in adults. We think that play might be the bedrock of the evolution of behaviors which allow the acquisition of 'novel' behaviors. Play behavior requires that the actions of an individual are decoupled from the need of fulfillment of daily life urgencies. Perhaps such a 'creative state' originally evolved to provide a 'test period' in which spontaneous behavior can take place (see [KG85]). Especially where individuals have to cope with a complex and dynamic environment, the search for new combinations of movements or new behaviors could be very advantageous in natural selection. Play is closely related to exploratory behavior (see [WV91]). This stresses the 'seeking for novelty' aspect in play behavior: known movements can be combined arbitrarily and might be applied to new situations resulting in new behavior patterns which increase the efficiency of the performance (improvement) or which allow the exploration of resources unknown up to this point (invention). In social play, an imitating agent is not searching for a specific solution, i.e. the imitated behavior need not be profitable to the imitator in the imitation situation, but rather it is tried out in the 'play mode'. Imitation should not be regarded as a kind of goal-directed optimization strategy, even if the acquisition of new skills will in the long run increase the fitness of the individual or the whole group. This play-aspect of imitation can also be applied to mimicry, Mitchell's first-level imitation (see section 4.1), since, in this case, the mechanisms of evolution itself 'play the game'.

So far this section has described why imitation is an interesting and challenging task for investigation, and not surprisingly, it has already attracted attention in robotics research, e.g. [HD94] and [Kun94]. Both approaches have in common that they focus on the 'non-social', functional aspect of imitation, i.e. implementing imitation behavior in order to let the model learn a specific task, i.e. maze negotiation of mobile robots ([HD94]) and imitation of a block assembly task which is performed by a human arm and copied by a robot manipulator ([Kun94]). While the latter approach uses the metaphor of imitation on a very general level and an 'engineering approach' to implement the desired functionality, the learning architecture proposed by Gillian Hayes and John Demiris ([HD94]) is deeply biologically motivated. The future potentials of imitative and possibly cross-modal learning as a technique for robots to learn from other robots is explained vividly in [HD94] by the idea that "'a robot which starts work when the sun rises could teach another to start work when it hears the birds singing". This might also be possible with other approaches in machine learning, but in nature imitation seems to be an effective tool which is used to learn from others, escpecially in the context of explicit teaching, without repeating the action numerous times, i.e. a kind of learning from very few or, in its extreme case, learning from one example. In this way imitation might be used as a promising tool for an implicit knowledge transfer between different kinds of robots or between robots and humans.

In both approaches mentioned above, imitation is not used as a 'social skill'; the imitator does not use imitation to 'get to know' the robot it is imitating. This is the main difference to the approach we are persuing. We do not use imitation only as a technique, an approach which may be absolutely sufficient in certain experimental conditions or industrial applications, but as a social skill. This does not exclude the possibility that our approach might also facilitate the functional aspect, as we will exemplify in the following: As long as the 'roles' in the imitation task are clear, e.g. one robot is always the model, the other the imitator, and no other potential models are present, then there seems to be no need for social relationships. But what should the imitator do if there is a group of robots with one imitator and several potential models? If all potential models behave in a similar way then the imitator can choose one model by chance and focus its attention on this model during the imitation process. If the potential models behave totally different, it might be highly advantageous to be able to distinguish between them, so that the model can be actively selected, at least after all candidates have been encountered once. In this way the imitator could choose models which have shown in the past to be 'good teachers'. And the situation becomes more complicated if mutual imitation (imitation 'games', no predefined roles) take place.

The next section gives arguments for the 'embodiment of social behavior' and describes our conception of 'body images'.

4.2 Body images

Data from psychology and child development show that the individual 'mental ontogeny' and the development of social contacts are strongly interrelated with the development of the physical body. Johnson ([Joh87]) argues that meaning, rationality and consciousness cannot be separated from bodily experiences and interactions with the environment. Rosenfield ([Ros93]) stresses that without reference to an individual with its unique point of view and without taking the body image of the self as a frame of reference, no perception, conscious awareness or memory would be possible "'It is not that my memories exist as stored images in my brain, conscious or unconscious; the act of memory is one of my relating to myself, or to others, or to past experiences, or to previously perceived stimuli. This is the very essence of memory: its self-referential base, its self-consciousness, ever evolving and ever changing, intrinsically dynamic and subjective." ([Ros93], p 8).

The relevance of body images for the development of 'self-concept', personality, social interactions and social management is impressively described in [vdV85]. Based on data from psychiatry, van der Velde presents a new concept of 'body image' (bi) referring to the own body and 'extraneous body image' (xbi) referring to the appearences and behavior of others. One line of argumentation in his work is very important for our 'social intelligence approach', namely that body images of others develop during interaction with the environment, especially through interaction with others. In a later phase of development these body images are the prerequisites for the development of the 'concept' of others. In the following paragraph some of the main principles of van der Velde's framework are presented.

According to van der Velde, all body images consist of two components. The physical component, representing a bodily feature or movement, is associated with the psychological meaning which comprises the thoughts and feelings underlying the physical component. Our perception of our own body or the bodies of others as 'entities' is the result of a conceptual composition of innumerable body images, partial representations of humans which we perceive for every part of the body. In ontogeny, the first functional body images which are formed during the first year of child development are body images of others, referring to the physical appearance and behavior of the mother. This takes place before the mother or any other thing in the outside world can be identified as a single and unique object (object permanence). At the same time, the child remembers the internal feeling states associated with the mother's appearance and actions. At the beginning, the infant might be capable of judging xbi's as pleasurable or displeasing. These internal feeling states are mnemonically fused with the xbi's. A mutual activation of xbi's and internal perceptions allows an indirect assess of internal emotional states. This enables a human infant to predict the consequences of the appearance of certain xbi's belonging to the mother and, if the mother's xbi's are consistent, pleasurable and presented with an adequate frequency, establish basic trust. This is the reason why, throughout life, the outer appearance and behavior of others are so closely interrelated with emotional states of ourselves. Therefore, xbi's are building blocks in the infant's mental construction of human objects and concepts of others, they are the basis for the development of attachment, basic trust and human interactions. Xbi's are gradually accumulated and 'become' human objects during the perceptual and cognitive development of the child. Van der Velde's conclusions imply that concepts of others result from the interaction with the social environment.

We should also note at this point that it is not the mere physical presence of a conspecific or group member which is important for social contacts. Direct bodily contacts play a major role, e.g. the "'cohesion of primate groups is maintained through time by social grooming"([Dun93]).

The next section describes the experimental setup which we have developed to study the development of artificial social intelligence, i.e. the experimental setup is used as a test-bed to investigate 'imitation' and 'body images'.

5 Experimental setup

To study the development of social relationships, we do not investigate imitation and the collection of body images in isolation. Instead, we use an ecological approach. The mechanisms are embedded in a scenario shown in fig. 2. This section is not intended to give a detailed description of our experimental approach. Technical details and experimental results will be presented in following papers. This section's goal is to describe the main ideas underlying the experimental approach, i.e. the 'test-bed'. This should outline how the theoretical considerations described in the previous chapters might be investigated in an experimental approach.

The subsections give answers to the questions 'Where should the robots interact?'(section 5.1), 'What kind of robots are used?' (section 5.2) and 'How should the robots interact?' (section 5.3 and section 5.4).

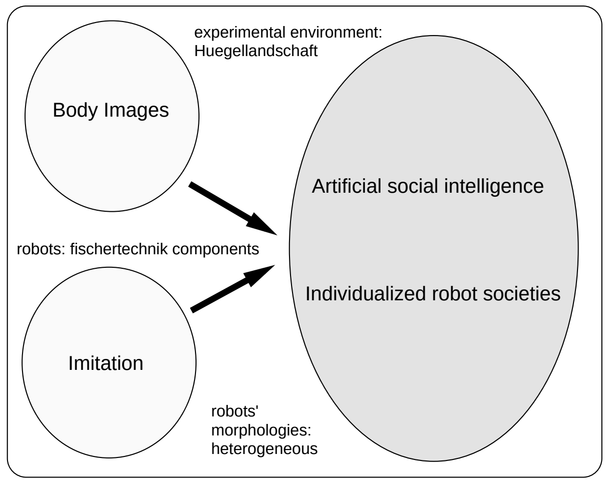

Figure 2: Scenario. Investigation of imitation and body images for studying the development of social intelligence.

5.1 The 'Huegellandschaft' habitat

Instead of a flat surface which is normally used for those robots which are designed to run in offices or laboratories, we use a hilly surface, the 'Huegellandschaft', which consists of smooth hills. Not all robots can go everywhere in the 'Huegellandschaft'. Some robots, e.g. those with weak motors, cannot climb any hill. The robots' velocity decreases with increasing inclination. The energy consumption is not only correlated with the horizontal distance covered, but also with the inclination. Therefore, by means of the 'Huegellandschaft', we can investigate activities which have a 'real' meaning to the 'life' of the robots. This is important for us in order to realize the concepts of the development of body images (see section 4.2). In addition, the activity of a robot depends on its expertise in keeping its energy level constant or at least in a certain interval which is a first step to include a sort of 'life-span' and, if a memory is used, a sort of simulated ontogeny (see section 3). In this way, it is not a mere 'simulation' of biological behaviors (see simulated mate-finding in [WD92], or food-collecting in [Mat94]). By introducing meaningful aspects of the world, the behaviors are not simulated but emerge out of the robot's 'self-interest' (see also [Ste94]).

For the following reasons, we hope that, beyond our specific approach, the 'Huegellandschaft' can also be an interesting test-bed for the study of other research issues in alife research.

- **Semi-natural habitat:** From a biological point of view, the 'Huegellandschaft' is a closer approximation to the complexity of natural surface than a plane. In landscapes which are not influenced by humans one can only rarely find mathematically exact planes (not unreasonably, animals do not use wheels). Since we do not work with walking machines, the 'Huegellandschaft' seems to be a reasonable compromise, a tradeoff between the natural complexity and the possibility to work with wheel-equipped robots.

- **Robot-environment interactions:** The inclination of the surface provides a direct way in which the surface provides 'meaning' to the robot while it is running on it. In order to survive in the 'Huegellandschaft', e.g. in order to avoid overturning, the robot must control the relation of its 'body' towards the environment. The robot has to 'know' its bodily characteristics, e.g. the overturning-angle. The individual variations can have large effects on the behavior of the robot, e.g. deciding whether it can climb a hill or not. The strategy to start an exploration in random directions can become highly energy consuming. The inclination of the surface can be seen as a smooth cost-function. If the robot cannot manage to climb a hill, the costs will become indefinitely high. This can be compared to the situation of a robot running on a plane cluttered with obstacles. But in this case we only have the binary distinction between space occupied by obstacles, on the one hand, and free-space on the other. In the 'Huegellandschaft', we have an continual transition between these two extreme cases where the robot can more or less 'enter' the obstacles. From a more abstract point of view, we see the 'Huegellandschaft' as a general case of a surface, while an empty plane or a plane cluttered with obstacles are two special cases. In the same way, the surface can also be modified to form a pipe (see fig. $3a$).

- **No mapping:** Since our approach focusses on robot-robot interactions, we exclude mapping behavior which is not essential for the survival of the robots in the 'Huegellandschaft'. But the robots are not lost in a 'terra incognita'. They can use local information provided by the inclination sensors, indicating the relationship between the environment and the robot's body axis. The inclination of the environment can be used as a local characterization of the environment. For instance, if a robot recognizes that an interesting 'device' (e.g. a recharging station) is placed on top of a steep hill, in case of low-energy, it can improve a random-search strategy by testing the steepness of the neigboring ground and choosing the steepest one. So, without global information, the local performance of the robots can be improved (see section 5.3).

- **Further ideas:** Fig. 3 also indicates what kind of behaviors would be possible if the surface is made sensitive to the weight of the robots, so that the robots can passively modify the form of the surface. The robots can also exploit this characteristic of the environment and can develop certain strategies, e.g. development of trails or building behavior. All these ideas suggest to investigate closer interactions and interdependencies of environment and robots only by virtue of the 'Huegellandschaft' without additional obstacles or object manipulation strategies.

PHOTO 1

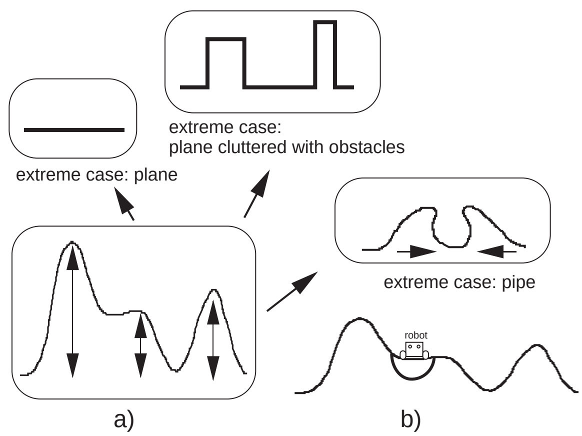

Figure 3: Robot-environment interactions in the Huegellandschaft. *a)* systematically varying the surface by the experimenter *b)* modification of the surface by moving robots, development of trails or 'building behavior'

5.2 Design of robots

An important principle in nature is variation. It is not a side-effect or a result of imprecise reproduction, but rather the basis for evolution. In nature, we cannot find two individuals with exactly the same outer appearance and behavior, even if their genetic code is identical. Simulations of communication and social interaction most often use 'clone-like' societies which can be produced much more easily (e.g. with copy & paste) than the modeling of 'individualized groups'. The same is true for most research on groups of robots. Using standard models of robots, no explict variation is used beyond unintentional variations in construction or different values for parameters. In contrast to these approaches we use a *heterogeneous* robot design.

Our experiments are performed with small robots which are built by using *fischertechnik* construction kits. We use robots with two driven front wheels and different kinds of sensors, e.g. light sensors (photo-transistors) and tilt sensors, contact switches and infrared sensors. Photo 1 shows one of our robots. The robots are at maximum 35 cm long and 30 cm wide. On board the robots are equipped with an energy supply (batteries) and a special on-board computer[7]. In order to have a very simple 'sensitive body surface', each robot has a belt around its body which is attached to contact sensors (see photo 2).

PHOTO 2

The robots are controlled using a behavior-oriented approach ([SV93]). Photo 3 shows our robots which are moving in the 'Huegellandschaft'. First tests have been conducted with behaviors like exploring the 'Huegellandschaft', moving up and down hills, controlling the body axis, keeping bodily contact, and distinguishing between moving and stationary objects (see [Dau95]).

PHOTO 3

5.3 Scenario

In order to investigate individual interactions between robots moving in the 'Huegellandschaft' we decided to pursue an incremental approach with three phases which are based upon each other, with an increasing complexity of the studied mechanisms.

- **Phase I: Sychronization and following of movements.**

  In this phase, the robots distinguish robots from other objects in the 'habitat' (the "'here is something like me"-distinction, see section 4.1), where robots are entities to whom they can establish physical contact and whose movements can be matched. The robots actively approach and seek for body contact with other robots, align to the main direction of movement and keep contact by synchronizing and following the movements of each other. These behaviors will help the robots to survive. The behaviors emerging at this stage can be classified as anonymous cooperation between heterogeneous robots. Figures 4 and 5 describe this idea.

- **Phase II: Recognition of others.**

  The robots are able to recognize individuals by storing information about other individuals (the "'this is robot XY which..." recognition mechanism, see section 4.1). As a first step towards collecting 'body images' of conspecifics during social interactions, they store information about movement characteristics of the

[7]Developed at VUB-AI Lab Brussels.

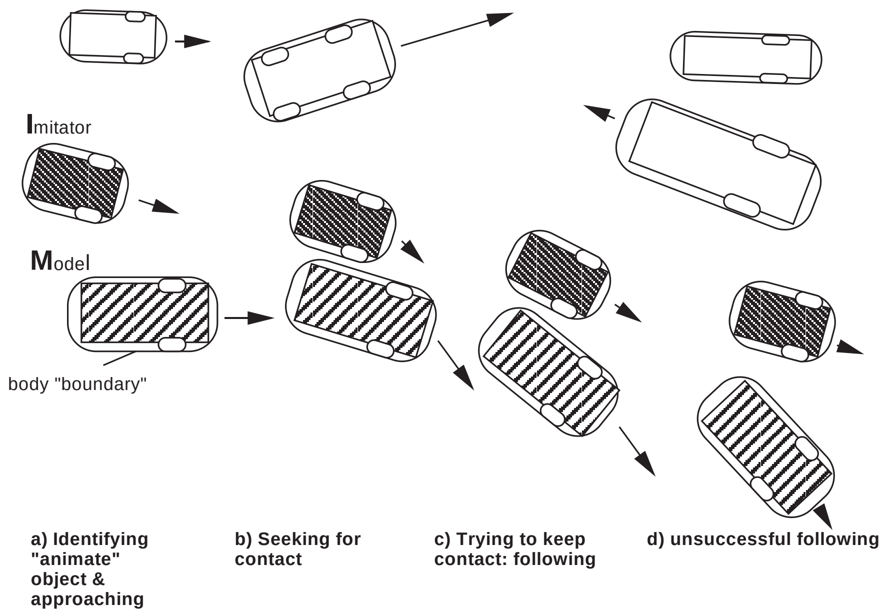

Figure 4: Starting point for phase I: initial stages of an individual, bodily contact between two robots.

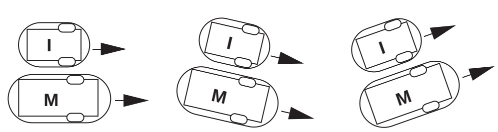

Figure 5: Successful 'communication' by synchronization of movements between the imitating *I-robot* and the *M-robot* as the model.

Learning strategies: "save energy"

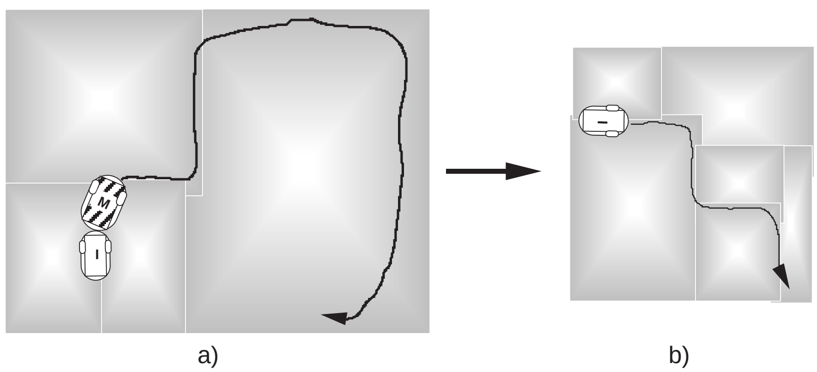

Figure 6: In its most elaborated case, not only movement patterns, but strategies which improve the fitness of the imitator which is running the ‘Huegellandschaft’ might be learnt. *a)* The imitator I follows the model M which runs in valleys. Robot I recognizes by means of self-observation, using information from its inclination sensors and batteries that the movements performed by the model are energy-saving. *b)* Robot I from now on can use the energy-saving strategy of moving in valleys.

other individual. The robots have to distinguish between robots with whom they can keep contact and those with whom they cannot. For instance, if the model is known to run too fast for the imitator, it would be energy-saving to avoid the model as soon as it is recognized. On the other hand, it might be advantageous in the Huegellandschaft to avoid or approach models which behave in an advantageous or disadvantegeous way, i.e. improving or worsening the 'fitness' of the imitator (see fig. 7). To achieve this, they should be able to attach meaning to the experiences they have gained with other robots. In the 'Huegellandschaft', meaning can be easily introduced by tracing the inclination of the surface which is correlated with the energy consumption of movements in order to distinguish 'pleasant' from 'unpleasant' experiences. In this way, while following the movements of others, an internal feedback is possible. This might lead to an interruption of the interaction, i.e. if the imitative movements are highly energy-consuming. The recognition or movement pattern is done in an 'egocentric' (self-centered) way, i.e. other's behavior is recognized by observing own movements. While they follow the movements of the other robot they 'observe' and classify their own body movements. We suppose that this self-observation might be one important factor in the learning of new movements. The importance of such a 'self-discovery' aspect in movement learning is also a central idea of the Feldenkrais method ([Nel89], [Ape92]). This self-observational aspect is explained in more detail in section 5.4. To summarize, in phase II the robots show a kind of 'behavior-oriented' recognition of conspecifics and 'personal' social relationships can develop, such as attachment or avoidance. This can be seen as the beginning of social management[8]. It can also provide the basis for the development of social structures (e.g. hierarchies). An example for the usefulness of the behavior-oriented recognition of 'conspecifics' in daily life applications might be the following situation: In an individualized group, the robots are able to detect and possibly separate 'conspecifics' within an unusual behavior, e.g. identify malfunctioning robots or robots with low energy level. This improves the survival capabilities of the whole group and might, for instance, protect the environment from destructive behavior of robots which are out of control. After the implementation of phase II, we hope that a long-term survival of the robots in the 'Huegellandschaft' is possible using the sychronizing and following of movements out of 'self-interest'. If the requirements of the daily 'struggle for survival' are fulfilled, in the third phase of our scenario, we want to investigate how the robots can effectively use their 'free time' and acquire new movement patterns by imitation.

- **Phase III: Movement learning by imitation.** In this phase, robots use their 'free time' (play-mode) to learn movements by imitating others. In phase

[8] In [Dau94], we propose one possible example for a symbiosis-like relationship between imitator and model, i.e. they help each other to overcome physical or sensory 'handicaps'.

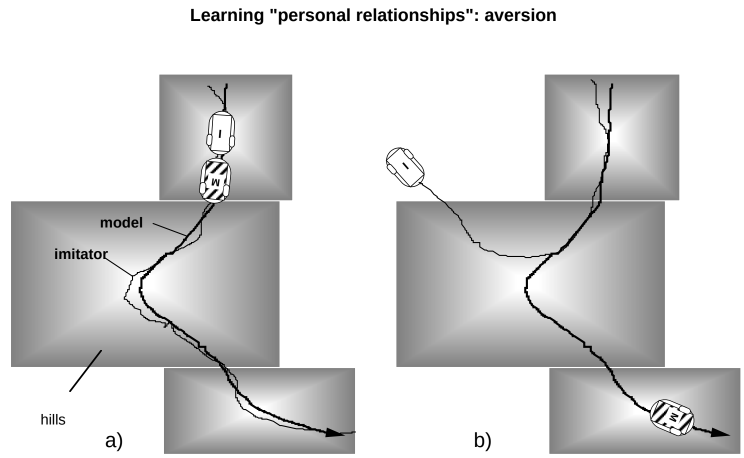

Figure 7: The development of attachment as an example of 'personal relationships'. *a)* The imitator *I* follows the model *M* which is running up and down hills, a highly energy-consuming behavior in the 'Huegellandschaft'. *b)* The next time robot *I* follows a model and recognizes that it is following robot *M*, robot *I* can interrupt the following behavior and avoid robot *M*.

III, robots may learn 'socially' relevant movements or new technical skills. Learning movements can refer to learning new movement patterns (see fig. 8) or, in its elaborated version, learning movement 'strategies' (see fig. 6). The latter will be the most complex form of imitative learning, since it requires, at least in a primitive way, a kind of 'reflective' ability. They acquire new knowledge which will enhance the fitness of the individual as well as the fitness of the whole group (see section 4.1 on imitation and play). At this stage it might also be possible to study the evolution of traditions.

The 'Huegellandschaft' is an interesting 'habitat' for imitative learning of movement patterns of strategies. Even if no internal map (using metrical or topological information) is used, the local feedback via inclination sensors provides local position information, e.g. using the information that the movement which is imitated starts and ends on hills. This does not protect the imitator from mistakes, but it is supposed to improve performance.

Biological literature on imitation usually discusses the imitation of *behaviors* rather than movement patterns. So far it is not clear how robots could *interpret* movement patterns as behaviors. This point has to be studied intensively in phase III. But we suppose that the attachment of meaning (interpretation of movements as positive or negative, see phase II) and the registration of the environmental interactions during the imitation process turn out to be

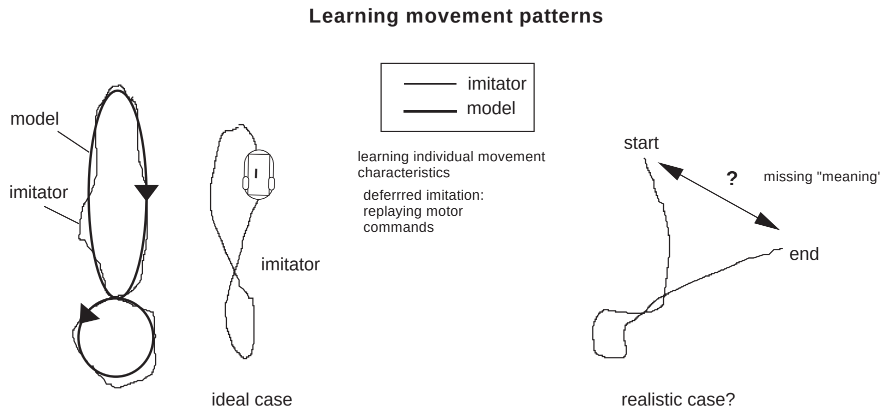

Figure 8: An example for learning movement patterns. The left side shows the trajectories of the imitating robot I which follows a model. In the ideal case, the imitating robot can more or less exactly replicate the movement pattern in the absence of the model, i.e. in a kind of 'deferred imitation'. On the right, a trajectory is shown which is supposed to represent the average case, i.e. small deviations during the self-observation process result in a qualitatively different movement pattern.

important factors.

5.4 Perceiving and Acting: Following and synchronization of movements

The left-hand side of fig. 10 shows the 'engineering' approach used for agents which should copy the movements of other agents. The 'teaching by showing' approaches of Yasuo Kuniyoshi, Hirochika Inoue and colleagues (e.g. [KII90], [KII92], [Kun94]) give examples. We call it an 'engineering' approach with respect to the main way which is pursued to produce the imitative action e.g. using an analytic approach of observing, segmenting and classifying the actions of the model, producing a high-level i.e. symbolic description of the observed movement and using this symbolic plan to control the movements of the imitator. In our point of view, this approach is restricted by the bottleneck of the need of mapping actions to and from a symbolic level to continuous real world actions. Assuming that humans and other animals are experts at imitation, we suppose that this approach will be as effective for real world applications, as humans who try to imitate the movements of each other by using only symbolic communication, namely language, via a telephone line. Because several levels of imitation (up to level 4 or 5 according to Mitchell's classification, see section 4.1) also occur in non-humans which are not supposed to have equally elaborated symbolic skills as humans do, symbolic descriptions are not expected to be a crucial prerequiste for the development of imitative skills. Young children are experts at imitation before they fully develop their symbolic skills. Since we are interested in the basic principles necesssary for imitation, hoping to reveal the basic underlying mechanisms and, by enhancement and adding of new ones, incrementally approaching higher levels of complexity, it is not adequate to start with an approach depending on a 'symbol level'.

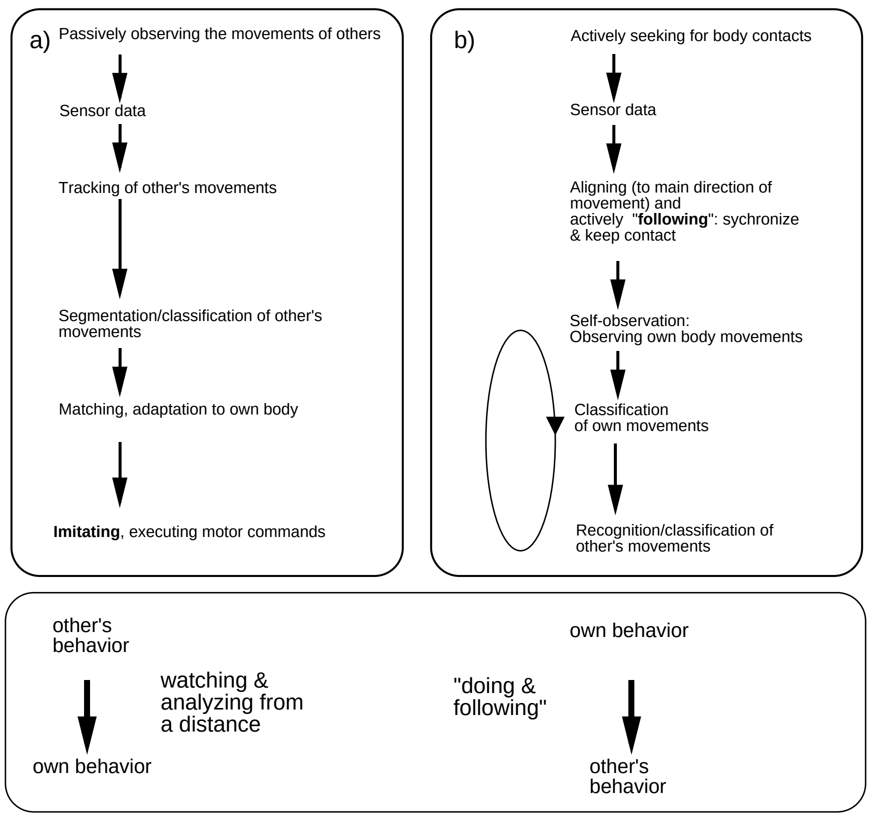

Figure 9: Two different ways of approaching imitation of movements. *a)* 'traditional approach', *b)* self-observational approach. The circle indicates a feedback loop, i.e. movement which are pleasurable or displeasing for the imitator can be used to cause interruption of the following behavior. For further explanation, see text.

In contrast to the approach discussed so far, we are using a different approach, shown on the right-hand side of fig. 10. The main idea is the self-observational aspect of the imitation action, i.e. the imitator does not 'watch and analyze the other, then acts itself' but rather uses a somehow opposite direction, namely 'acts and follows, then analyzes its own movements. While following the first strategy, the action of the imitator only takes place after the movements of the other have been analyzed and transferred to the 'own body', namely the movement capabilities and morphology of the imitator, the second strategy starts with movement from the beginning, following and synchronizing the own movements with the model's movements (with reactive, local rules). The analysis providing information about the model's movements is done during the imitation process through 'introspection', i.e. through analyzing the imitator's own movements. According to Mitchell's classification (see section 4.1), we are investigating imitation of movements at level 3 where the robots strive to achieve correspondence with the model.

But since the analysis, i.e. recognition of movements, still has to be done, are there any advantages of this approach? We just want to mention four arguments, whereby the first two of them strongly refer to the 'meaning of the movement', e.g. when learning an ecologically or socially relevant movement. (1) Even if the imitator is not successful in recognizing the model by simply 'keeping' contact, it might be advantageous, if the execution or result of the movements is of any benefit to the robots. (2) The 'meaning' of the movement need not be deduced from the observation of the model. It can be experienced by the imitator itself, e.g. by giving sensor data from the external environment or the internal states. For instance, if the movement of the model is highly energy-consuming, this can be detected through self-observation. (3) The interpretation of the model's movements and adaptation to the movements of the imitator is facilitated: Since the imitator only interprets the imitative movement with respect to its own 'knowledge' and 'capabilites', it needs not know what this movement means to the model. If model and imitator are robots of the same 'species' inhabiting the same environment, it is highly probable that their interpretation is similar too. (4) There is no communication by explicitly exchanging signals, although communication takes place. But it is more a communication of movement sequences (the interpretation depending on the 'recipient', e.g. the imitator) than a communication of symbols. The imitator and the model need not have the same 'language' or mode of communication which is important for imitation between different kind of robots (see section 4.1 about cross-modal learning) and for robots imitating humans. Since our approach to the study of imitation does not depend on explicit symbolic communication, it is applicable to different kinds of robots (see heterogeneous robots, section 5.2).

Although our approach argues against the traditional one-way direction from perception to action, we do not promote a crude 'act first, perceive later' strategy. Instead, the imitator, while imitating the model, should trace and control continuously so that it does not come into dangerous situations, e.g. moving up a hill until it overturns. The conditions which require that the imitating process be interrupted do not only serve to maintain the survival of the imitator, but can also be used as an important source of information about the behavior of the model. In sum, we focus on the interdependencies between (social-) interaction, on the one hand, and perception, on the other. With respect to the relationship between action and perception, the idea presented in [PS94] of an ongoing co-ordination between perception and action is closely related to our approach. Pfeifer ([PS94]) pointed out how this different point of view can be interpreted as a design heuristic, e.g. when choosing appropriate sensors for a specific 'task'. In the same way, we suppose that starting with considerations about social interactions is important for decisions concerning the design of robots, whereby 'design' refers to control and construction principles. For instance, if bodily contacts play a crucial role in the interaction process, and especially if the body shape is used as an individual characteristic of a robot, then we have to think carefully about tactile sensors and the shape and flexibility of the body surface. In sum, we think that the shapes and bodily characteristics of robots are too often dominated by technical constraints (e.g. availability of components) or adaptation to the non-social environment (see [Tod93]). Instead, the robot's interactions both with the 'inanimate' environment and its social interaction partners should rather influence the design process.

The rest of the paper is dealing with general considerations about the relevance of the approach presented in this paper for autonomous robot research. Some ideas about possible relationships between robot and human societies are presented. Societies of autonomous robots do currently not exist, therefore, these considerations have to be speculative and vague. But since the scenarios we have in mind for future robot societies will influence the research activities on these issues, we outline some aspects which might be important to discuss and investigate empirically in more detail.

6 Autonomous and socially bonded robots in real world applications

The previous sections of this paper intensively described the biological motivation to study artificial social intelligence. But to build social robots in order to understand something about the mechanisms and the development of social and non-social intelligence is not our only motivation in studying these phenomena. We also hope that it will contribute to application oriented research on autonomus robots (see section 2). The answer to the question 'do autonomous robots need social intelligence?' is not simply 'yes' or 'no', but 'it depends...'. It depends on the application field, on the goals and expectations of humans and on the intensity of contacts of both robot designers and robot users with robots. The use of principles which are important to biological or artificial 'creatures' is only sensible if it promises to improve the performance of the robots in the specific application field. If technical solutions are absolutely sufficient, it would be over-engineered to incorporate alife principles. So, under what circumstances does it seem to be useful to follow an alife approach, i.e. to have robots with 'social intelligence'?

- **The need for group behavior.** If a single robot suffices to fulfill a specific task without explicitly coordinating its actions with humans or artifacts, i.e. if it is never encountered in any kind of social interactions, then there seems to be no reason why such a robot should develop 'social intelligence'. The situation is different if robots have to cooperate in order to fulfill a specific task. Either the task at hand technically demands for more than one robot (e.g. transportation of heavy objects or time constraints), or groups should explicitly cooperate (e.g. divide labour, help each other out of deadlocks). While, in the former example, the specific problem might be solved only by communication via the environment, the latter tasks indicate the need for communication, e.g. the robot exchange information, either implicitly (just broadcasting signals) or explicitly with one-to-one communication. Our suggestion that the recognition of 'conspecifics' might enhance the performance of groups of robots is supported by Toshio Fukuda's results on a distributed autonomous robotic system (CEBOT), showing that recognition of others as group members and recognition of their behavior can be used to optimize group behavior (see [FIUA94]).

- **The need for 'individualized robots'.** Nature gives us many successful models of cooperation between natural agents. In the case of the transporting task mentioned above, we can think of the metaphor of social insects. This metaphor holds in two aspects: (1) concerning the self-organizational development of complex global patterns out of local rules of a distributed system and (2) the 'strategy' of nature to bet on number instead of quality, i.e. the agents themselves are relatively simple (less energy invested). The number of individuals is the most relevant aspect. Consequently, not the survival of the individual but of the colony (a super-organism) is selected, and the welfare of the single agent (unless it has a crucial role in the organization of the whole system) does not count much. So, with respect to robots, if the designer obeys the strategy to use a great number of homogeneous, interchangeable, 'cheap & simple' robots which should interact, this might suffice in many applications. If one robot gets lost (gets stuck in deadlocks or breaks down), it can be replaced easily. The situation is totally different if much money and technology is invested in the production of a single robot. We might think of a team of 'expert robots' with different functionalities. A team of sewage robots inspecting, mapping, repairing and controlling a sewage system is a good example. This corresponds to the second main 'strategy' of nature to invest much in the single individual, producing only a small number and trying to ensure the survival of the individual by providing either good environmental and/or social conditions.

- **The need for social communication.** Individuals, e.g. in the expert teams mentioned above, would highly benefit from the ability to help each other out of dangerous situations. But this requires the ability to recognize each other as individuals (e.g. detect when the other one changes its behavior due to decreasing battery level) and have a 'social bonding', e.g. an 'interest' to help each other. Additionally, in different situations they have to agree upon the responsibility of each other in order not to hinder each other. The possibility to attribute goals, knowledge or any kind of individual characteristics to other robots would improve the performance of the whole group. The case of an expert team in a sewage system shows another argument for the benefit of 'social intelligent' robots, namely not only the individual robot has to care about its survival, but there is a strong interest in the survival of the whole group ('they have to come back'), which has to be managed *autonomously*.

- **The need for autonomy.** Any kind of communication or social interaction is useless if the behavior of the robots can be alternatively controlled by human operators or via an overall central control system. If transportation vehicles in a store house can be controlled effectively via radio link by a central computer, there is no need for the robots to interact explicitly at all. In this case, the 'social intelligence' is located in the planning and scheduling routines of the central computer. But groups of robots working in highly unpredictable outdoor environments, where a central control is either not possible or not efficient enough, need to act autonomously.

- **A question of number.** A great amount of knowledge about individuals is not possible for an arbitrary number of individuals. 'Individualized' natural societies do not consist of arbitrary large groups. Even humans who are inhabitants of large towns and who daily encounter a great number of people do not establish intense personal relationships with all these people. Instead, "'...there is a cognitive limit to the number of individuals with whom any one person can maintain stable relationships". This number is about 150 ([Dun93]). In much larger groups, Dunbar suggests, we use our language as a technique to categorize people and to structure relationships hierarchically, while intense relationships can only be maintained to 10-12 people at any one time. Artificial 'brains' might be able to maintain larger numbers of intense relationships (which will probably be different and not as intense as among humans), but there might also be a threshold in the number of group members. It would be highly interesting to perform systematic investigations with reals robots, finding out the most adaptive group size depending on environment, group behavior and other constraints.

We hope that the arguments and examples given above outline why we think that the alife approach and especially our field of research, namely the development of social intelligence, might be highly useful in those autonomous robot applications where groups of 'individualized' robots must autonomously manage to ensure the survival of the whole group, involving social interactions and communication. Since for all these aspects, nature shows us highly efficient examples, we think that robotics research could profit from alife research, trying to find out the 'basic principles' applicable to both biological systems and artifacts.

Up to now in this section we have talked about humans only in the context of designers. But humans are also users and 'fellow-creatures' of robots, i.e. even if the single human does not possess an autonomous robot, it seems that one cannot avoid to encounter them in future daily life. The following points discuss the relationships between humans and robots and between human and robot societies.

- **Robots as effective machines.** We do not have to care about social capabilities of robots which we use as more or less intelligent machines as in the case with most robots existing today. They fulfill specified tasks – building cars, moving the lawn and so on. Future service robots will have more elaborated skills, e.g. used as intelligent vehicles (wheel chair), serving out meals in hospitals and so on. First realizations are already available. And in the same way, as we usually do not speak to our vacuum cleaner (as we do to our pets), people do not question whether groups of vaccum cleaner robots should be social or not or what kind of social structure they should have. People would probably prefer to treat them as machines. So the question 'anonymous or individualized society' is reduced to whether this will increase or decrease the robots' functionality.

- **Robots outside the human society.** The aspect of autonomy is highly important, especially in those 'habitats' where robots should work in dangerous, hostile, or inaccessible environments, fulfilling certain tasks instead of humans, e.g. inspecting sewage systems, collecting rocks on other planets, or (micro-robots) navigating in blood vessels. Many approaches to autonomous robots, and alife too, envision these kinds of applications. The robots must be able to communicate with humans, but usually only in distinct phases, and not for cooperation. The habitat which these robots 'inhabit' is spatially, temporally or dimensionally separated from human daily life. Consequently robot designers would use any kind of behavior, social or non-social, as long as it improves the performance of the robots. Those kinds of robot societies might seem to humans like 'a different kind of species', of strange and 'abnormal' behavior. But since they do not belong to our daily life, people do not care much about them.

- **Robots as social interaction partners.** Applications developed for future service robots often require the need for intensive communication and interaction with humans. A communication process is not only an information exchange process, but it is always embedded in a personal relationship between agents. For instance, for the acceptance of robots working as a filling station attendant, it can become necessary to be able to recognize individual customers who come regularly to the same filling station and expect to be treated individually. Consequently, robots working in service applications (e.g. in information bureaus, reception offices or any kind of shop-assistant position) have to face the problem that the robots should be accepted by humans as communication partners. If people are used to and enjoy interacting to a human and not to a machine, they might not accept the robot. Instead, the robots should be equipped with at least some social skills of human interaction partners. This tendency also seems to influence the hardware design (see attempts to build humanoid bodies, e.g. [BS93], [IS94]).

Up to this point, the arguments still holds that 'little by little people will get used to a special kind of interaction with machines', similar to the fact that many people got used to programming their video recorder (even if they hate doing so). But mental experiences are collected during life-time and contribute to the development of an individual character (see section 3). We cannot switch off this characteristic, choosing good examples and excluding those interactions which we do not want to have an influence upon us. Therefore intensive interactions to artifacts can influence our conceptions of interaction and communication in general. This effect is much more powerful for young people or children, when the development of 'social intelligence' is just at the beginning. We do not want to discuss this point here in great detail, but we want to mention a similar discussion in another technical domain working on the creation of artifacts, namely virtual reality: Most of Valerie E. Stone's ([Sto93]) arguments as to why and how non-social VR environments (where the user is only interacting with simulations of people) might influence children's social development can also be applied to interactions of humans with 'real artifacts', namely robots. Having this in mind, we might rethink the idea of using service robots as playmates for children ([IS94]) as soon as the robots have enough technical skills. If future robots should be integrated in the complex social system of our 'primate society', robots must acquire 'social expertise' in order to communicate and be accepted by humans.

- **Robots as competitive species?** We do not want to discuss in detail the scenarios which suggest a future 'species competition' between robot and human societies. All comments on this have to be higly speculative and vague. We only want to mention that this need not be a logical consequence. The alternative might be the *integration* of intelligent artifacts in natural societies. If artifacts are 'socially bonded' to humans, they must not necessarily be 'interested' in forming an alliance against mankind even if they probably will (one day) perform many functions remarkably better than humans. In science fiction literature, the videogame market and so on the human-robot competition is a very popular idea (the most popular idea seems to be human-human aggression). But although inter- and intraspecifc aggression and competition plays an important role in the life of all species, the most 'intelligent' species live in complex social systems with a high degree of stability during long periods of time. The idea that robots (worldwide) recognize each other as belonging to the same 'species', and would agree on attacking or suppressing the 'human species' seems to us as speculative as assuming this kind of behavior for birds. Instead, human families keeping dogs or cats are our alternative suggestion for a model of (most often) successful multi-species societies.

7 Future development of autonomous robots

We distinguish three different ways in approaching the common goal of 'creation of intelligent artifacts'[9]. The 'robotics approach' usually aims at building perfect 'machines', i.e. with high expertise at fulfilling a specific task, e.g. fast and precisely navigating in a ware-house (e.g. using 'traditional' approaches of ai, expert systems, machine learning...). The second one is the 'bionics' approach: taking principles and details from biological systems and transfering them to technical domains, i.e. artifacts (e.g. [WEP93]), even if the biological system has evolved in a different 'ecological' context and with different morphological constraints (e.g. size or weight). Many neural net approaches also take the principle of biological neural information processing simply as a 'tool' and source of inspiration, used as an optimization method and applied to various contexts. The third one is the synthetic alife approach inspired by Langton's ideas [Lan89], 'creating' artifacts, and taking seriously the context ('habitat') where the 'creatures' evolve. According to Langton alife "'complements the traditional biological sciences concerned with the analysis of living organisms by attempting to synthesize life-like behaviors within computers and other artificial media." These artifacts are not basically designed to fulfill a task prespecified by humans. The dream of many alife researchers is to build autonomous artifacts which 'survive' in their 'habitat' without external control. The dream would be fulfilled even if these beings were to behave in a way not intended originally by the designer, i.e. behaving 'foolish' in the eye of the human observer, e.g. interfering with humans interests (we only have to think of 'rat-like' artificats interacting with similar complexity with the animate and inanimate environment).

These different approaches are highly overlapping and have no diverging 'lines' of development. Important results in the technical domains, e.g. concerning movement capabilities (walking, climbing) or sensor equipment will be taken over with enthusiasm by alife researchers, since they extend the experimental possibilities. Moreover, bionics and especially alife themselves are highly interdisciplinary approaches and cannot exist without cooperation and interaction with other disciplines.

The research direction of 'artificial social intelligence' presented in this paper

[9] Many research areas contribute to the construction of autonomous robots but we just want to mention three areas which seem to us most important.

is also highly related to biological, psychological and other natural science research areas. A common framework in the study of principles of social intelligence might be a long-term, challenging task for basic and interdisciplinary research, in its ideal case derived from parallel studies in biological and technical sciences. As it is pointed out by Whiten and Byrne ([WB88]) "'...we need to find out *how* and in what forms intelligence works in the context of social interaction." The 'intellect' of most present day autonomous robots does not exceed the insect level, but long-term projects (see section 2) are already envisioning robots which should be normal parts of our everyday life. Future societies might consist of people who interact and communicate in their everyday life with other people and physical and virtual artifacts. Therefore, we should start early to think about (1) artificial social intelligence for autonomous robots, and (2) about robot societies and a symbiosis, and possible integration of robots into human societies. This might be the basis for a social and cultural 'coevolution' process between robots and humans, since "'Through the emergence of new tools, we come to a changing awareness of human nature and human action, which in turn leads to new technological development" ([WF86], p 163).

Acknowledgements

Thanks to all those who helped me with fruitful discussions and comments on my work. I am especially grateful to Thomas Christaller for giving me the chance to investigate those research issues I am highly interested in. I also thank him for many valuable discussions, suggestions and comments on my work and for encouraging me to develop the ideas which have gone into this paper.

References

[Ape92] U. Apel. The Feldenkrais method: awareness through movement. *WHO Regional Publications/European Series*, 44:324–327, 1992.

[Aro94] Elliot Aronson. *The social animal.* W.H. Freeman and Company, New York, 1994.

[Bat94] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[BC92] R. J. Blumenschine and J. A. Cavallo. Scavenging and human evolution. *Scientific American*, pages 70–76, October 1992.

[Bro91] R. A. Brooks. Intelligence without reason. Memo 1293, MIT, 1991.

[BS93] R. A. Brooks and L. A. Stein. Building brains for bodies. Memo 1439, MIT, 1993.

[Bul83] D. Bullock. Seeking relations between cognitive and social-interactive transitions. In K. w. Fischer, editor, *Levels and transitions in children's development: new directions for child development*, chapter 7. Jossey-Bass Inc., 1983.

[Bul87] D. Bullock. Socializing the theory of intellectual development. In M. chapman and R. A. Dixon, editors, *Meaning and the growth of understanding: Wittgenstein's significance for developmental psychology*, chapter 11. Springer-Verlag, 1987.

[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.

[CM53] M. R. A. Chance and A. P. Mead. Social behaviour and primate evolution. *Symp. soc. exp. Biol. VII (Evolution)*, pages 395–439, 1953.

[CS92] D. L. Cheney and R. M. Seyfarth. Précis of how monkeys see the world. *Behavioral and Brain Sciences*, 15:135–182, 1992.

[Cyp93] A. Cypher, editor. *Watch what I do: Programming by demonstration*. MIT Press, 1993.

[Dau94] Kerstin Dautenhahn. Trying to imitate – a step towards releasing robots from social isolation. In P. Gaussier and J.-D. Nicoud, editors, *Proc. From Perception to Action Conference, Lausanne, Switzerland*, pages 290–301. IEEE Computer Society Press, 1994.

[Dau95] Kerstin Dautenhahn. Robots keeping bodily contact – following of movements as a step towards social interactions. Abstracts Book of ECAL'95, 3rd European Conference on Artificial Life, Granada, Spain, 1995.

[DC95] Kerstin Dautenhahn and Thomas Christaller. Remembering, rehearsal and empathy: Inspirations from a dynamic systems point of view. AISB Workshop 'Reaching for Mind: Foundations of Cognitive Science', Workshop Notes, 1995.

[DGF+91] J. L. Deneubourg, S. Goss, N. Franks, A. Sendova-Franks, C. Detrain, and L. Chrétien. The dynamics of collective sorting: robot-like ants and ant-like robots. In J. A. Meyer and S. W. Wilson, editors, *From animals to animats, Proc. of the First International Conference on simulation of adaptive behavior*, pages 356–363, 1991.

[Dun93] R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–735, 1993.

[FhG94] FhG. Serviceroboter – ein Beitrag zur Innovation im Dienstleistungswesen. Fraunhofer-Institut für Produktionstechnik und Automatisierung (IPA), Stuttgart, 1994.

[Fis93] J. Fischman. New clues surface about the making of the mind. *Science*, 262:1517, 1993.

[FIUA94] T. Fukuda, G. Iritani, T. Ueyama, and F. Arain. Self-organizing robotic systems, organisation and evolution of group behavior in cellular robotic system. In P. Gaussier and J.-D. Nicoud, editors, *Proc. From Perception to Action Conference, Lausanne, Switzerland*, pages 24–35. IEEE Computer Society Press, 1994.

[Ge94] B. Grosz and R. Davis (eds.). A report to arpa on twenty-first century intelligent systems. *AI Magazine*, 15(3):10–20, 1994.

[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.

[Hum76] N. Humphrey. The social function of intellect. In P. P. G. Bateson and R. A. Hinde, editors, *Growing points in ethology*, pages 303–317. Cambridge University Press, 1976.

[Hum84] N. Humphrey. *Consciousness regained.* Oxford University Press, 1984.

[IS94] H. Inoue and T. Sato. Humanoid autonomous system. RWC Technical Report, TR-94001, Special Issue, 1994.

[Joh87] M. Johnson. *The body in the mind.* University of Chicago Press, Chicago, London, 1987.

[Jol66] A. Jolly. Lemur social behavior and primate intelligence. *Science*, 153:501–506, 1966.

[Kat91] I. Kato. Homini-robotism. In *Fifth Internat. Conference on Advanced Robotics*, volume 1, pages 1–5, 1991.

[KG85] H. Kummer and J. Goodall. Conditions of innovative behaviour in primates. *Phil. Trans. R. Soc. Lond. B 308*, pages 203–214, 1985.

[KII90] Y. Kuniyoshi, H. Inoue, and M. Inaba. Design and implementation of a system that generates assembly programs from visual recognition of human action sequences. In *Proc. IEEE International Workshop on Intelligent Robots and Systems IROS '90*, pages 567–574, 1990.

[KII92] Y. Kuniyoshi, M. Inaba, and H. Inoue. Seeing, understanding and doing human task. In *Proc. 1992 IEEE International Conference on Robotics and Automation*, pages 1–9, 1992.

[Kun94] Y. Kuniyoshi. The science of imitation – towards physically and socially grounded intelligence. RWC Technical Report, TR-94001, Special Issue, 1994.

[KZ94] C. R. Kube and H. Z. Zhang. Collective robotics: from social insects to robots. *Adaptive Behavior*, 2(2):189–218, 1994.

[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.

[Mat94] M. J. Mataric. Learning to behave socially. In *From animals to animats, Proc. of the Third International Conference on Simulation of Adaptive Behavior*, 1994.

[Mat95] Maja J. Mataric. Issues and approaches in design of collective autonomous agents. *Robotics and Autonomous Systems*, 16:321–331, 1995.

[MG93] A. Meltzoff and A. Gopnik. The role of imitation in understanding persons and developing a theory of mind. In S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, editors, *Understanding other minds*, pages 335–366. Oxford University Press, 1993.

[Mil93] K. Milton. Diet and primate evolution. *Scientific American*, pages 70–77, 1993.

[Mit87] R. W. Mitchell. A comparative-developmental approach to understanding imitation. P. P. G. Bateson (ed.): Perspectives in Ethology 7, 183-215, 1987.

[MM92] A. N. Meltzoff and M. K. Moore. Early imitation within a functional framework: the importance of person identity, movement, and development. *Infant Behavior and Development*, 15:479–505, 1992.

[Moo92] B. R. Moore. Avian movement imitation and a new form of mimicry: tracing the evolution of a complex form of learning. *Behaviour*, 122:231–263, 1992.

[MPM+94] P. Marchal, C. Piguet, D. Mange, A. Stauffer, and S. Durand. Embryological development on silicon. In R. A. Brooks and P. Maes, editors, *Artificial Life IV, Proc. of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pages 365–370, 1994.

[MV87] H. Maturana and F. Varela. *The tree of knowledge: the biological roots of human understanding.* New Science Library, Boston, 1987.

[Nel89] S. H. Nelson. Playing with the entire self: the feldenkrais method and musicians. *Seminars in Neurology*, 9(2):97–104, 1989.

[Ots93] N. Otsu. Toward flexible intelligence: Miti's new program of real world computing. In *Proc. IJCAI-93*, pages 786–791, 1993.

[PS94] R. Pfeifer and C. Scheier. From perception to action: the right direction? In P. Gaussier and J.-D. Nicoud, editors, *Proc. From Perception to Action Conference, Lausanne, Switzerland*, pages 1–11. IEEE Computer Society Press, 1994.

[Rie94] Doug Riecken. Introduction to special issue on intelligent agents. *Communications of the ACM*, 37(7):18–21, 1994.

[Ros93] I. Rosenfield. *The strange, familiar, and forgotten. An anatomy of consciousness.* Vintage Books, 1993.

[Ste94] Luc Steels. A case study in the behavior-oriented design of autonomous agents. In D. Cliff, P. Husbands, J.-A. Meyer, and S. W. Wilson, editors, *From Animals to Animats 3, Proc. of the Third International Conference on Simulation of Adaptive Behavior*, pages 445–452. IEEE Computer Society Press, 1994.

[Sto93] V. E. Stone. Social interaction and social development in virtual environments. *Presence*, 2(2):153–161, 1993.

[SV93] Luc Steels and Filip Vertommen. Emergent behavior. A case study for wall following. VUB AI Lab Memo, 1993.

[TGGD91] G. Theraulaz, S. Goss, J. Gervet, and L. J. Deneubourg. Task differentiation in polistes wasp colonies: a model for self-organizing groups of robots. In J. A. Meyer and S. W. Wilson, editors, *From animals to animats, Proc. of the First International Conference on simulation of adaptive behavior*, pages 346–355, 1991.

[Tho63] W. H. Thorpe. *Learning and instinct in animals.* London: Methuen, 1963.

[TM93] S. B. Thrun and T. M. Mitchell. Lifelong robot learning. Technical Report, IAI-TR-93-7, Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Informatik III, 1993.

[Tod93] D. J. Todd. Mobile robots – the lessons from nature. In P. Dario, G. Sandini, and P. Aebischer, editors, *Robots and biological systems: towards a new bionics?*, pages 193–206. Springer-Verlag, 1993.

[vdV85] C. D. van der Velde. Body images of one's self and of others: developmental and clinical significance. *American Journal of Psychiatry*, 142(5):527–537, 1985.

[Wat95] Stuart Watt. The naive psychology manifesto. The Open University, Knowledge Media Institute, number KMI-TR-12, 1995.

[WB88] A. Whiten and R. W. Byrne. The machiavellian intelligence hypotheses: editorial. In R. W. Byrne and A. Whiten, editors, *Machiavellian intelligence*, chapter 1. Clarendon Press, 1988.

[WD92] G. M. Werner and M. G. Dyer. Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, and J. D. Framer, editors, *Proc. of the Second Artificial Life Workshop*, pages 659–687, 1992.

[WEP93] H. J. Weidemann, J. Eltze, and F. Pfeiffer. Leg design based on biological principles. In *Proc. 1993 IEEE International Conference on Robotics and Automation*, pages 352–358, 1993.

[WF86] T. Winograd and F. Flores. *Understanding computers and cognition: a new foundation for design.* Ablex Publishing Corporation, New Jersey, 1986.

[WMMH92] I. H. Witten, B. A. MacDonald, D. L. Maulsby, and R. Heise. Programming by example: the human face of ai. *AI and Society*, 6:166–185, 1992.

[WV91] D. G. M. Woodgush and K. Vestergaard. The seeking for novelty and its relation to play. *Animal Behaviour*, 42:599–606, 1991.