Embodied Cognition in Animals and Artifacts

Kerstin Dautenhahn
Vrije Universiteit Brussel (VUB)
Artificial Intelligence Laboratory
Pleinlaan 2, 1050 Brussels, Belgium
kerstin@arti.vub.ac.be

Abstract

Based on experimental findings from the study of human cognition we discuss the concept of embodiment for artifacts. We argue that embodiment is linked to a *concept* of a body and is not necessarily given when running a control program on robot hardware. For this purpose we introduce the concept of 'body image'. Additionally, we stress the *individual* characteristics of an embodied cognitive system, as well as its *social embeddedness*. We then discuss how research on embodiment for artifacts might contribute to bridge the gap between phenomenological understanding and computationalistic approaches which traditionally dominate computer science and cognitive science. We outline a framework of the *physical-psychological state-space* which changes dynamically in a self-modifying way as a holistic approach towards embodied human and artificial cognition. This metapher is meant as a basis for discussion towards a common framework for cognitive architectures comprising natural and artificial systems. In order to stress the importance of a dynamic memory we introduce the concept of an 'autobiographical agent'.

Introduction

The ideas presented in this paper derive from our actual work on social intelligence in groups of autonomous, physical agents (Dautenhahn 1995). We am mostly interested in how, beyond simple behaviors which are necessary for the survival of a robot in an ecosystem, complex (socially) intelligent systems can develop.

For several years the New Artificial Intelligence approach ((Brooks 1991), (Pfeifer 1995)) put emphasis on the study of embodied agents, namely robots, interacting in a real environment. Recently new kind of projects have appeared, trying the scaling-up from relatively simple, mostly reactive systems towards cognitive systems, e.g. including language skills (Steels 1996) or trying to achieve the developmental cognitive state of a two-year old child (Brooks & Stein 1993). But instead of the common agreement that cognition can only develop in a body and through its interactions with the environment, the concept of *embodiment* has mostly been reduced to the fact of possessing and controlling a body, i.e having a internal control program which uses the body as an actuator device. However, there is much evidence to support the assumption that cognitive capabilities are only possible through the interaction of body and mind, i.e. that the body is not simply *used* by the mind, but that there is a co-development and mutual shaping of cognitive abilities on the one hand and bodily skills and experiences on the other hand. The body is not a fixed and pregiven 'actuator device', but it is a dynamic and ontogenetically evolving entity. Evidence from brain sciences and neuropsychology stress the mutual dependence of body and mind/personality. In the following sections we outline some aspects which we consider important to the design of embodied *cognitive robots* motivated by evidence from research in natural sciences. The concept of body image is central to our view. Body image refers to a dynamic construction and does not replace the mind in the body metapher by its contrary. The constructivist and dynamic nature of our conception of embodiment is described in more detail in the following sections. We sketch the concept of a physical-psychological state-space as a possible common framework for embodied cognitive architectures (details are described in (Dautenhahn & Christaller 1996)). This concept is not meant as a technical specification for the construction of robots. Instead, we intend to give some insights into general design guidelines for robots. For instance: if we take the holistic approach towards embodiment seriously then this would mean that the modification of a robot (adding additional wheels, manipulators or sensors) should influence not only its bodily skills but also its cognition. The general methodology of an incremental and modular construction of a robot has to face the underlying assumption of a modular nature of cognition!

Bodily Experiences in Real and Simulated Worlds

If we run a control program on a robot hardware platform (with real sensors, actuators, and in a real environment) why should this necessarily fulfill the embodiment criterion? Of course, the behavior of a real world robot is usually different from that of a simulated robot-environment system. But is this not mainly a matter of whether the complexity of nature and its laws can be captured by a computer program? There have been various discussions on the 'simulation versus real world' debate which should not be reviewed here (e.g. (Brooks & Mataric 1994)). In our view, the debate itself is not very useful in the sense of trying to find the 'right answer' to the question of what environment to chose for artificial intelligence or artificial life research. Humans are from their early childhood on experts at taking various abstract or fictional things for real entities (comprising comic, television-, or video game characters, football teams as well as political theories and religion). What counts is to our view the question of what is real to the embodied mind of an individual person. Following the argumentation of radical constructivism (e.g. (Roth 1994)) it should be more useful to discuss about the individual's constructed conception of reality (*Wirklichkeit*) than about an objective reality. The *meaning*, and not the technological basis is central. In this way, experiences which are important to the life of an individual should be taken seriously, no matter if they originate in real, simulated, virtual or fictional experiences. But, and this leads us back to the topic of this article, why is it nevertheless possible for humans to distinguish easily between real experiences and those watched on television, or read in a novel? Our hypothesis is that humans use their body actively in order to *test* the environment. Bodily interacting with the world is the easiest way to learn about the world, because it directly provides meaning, context, the 'right' perspective, and sensory feedback. Moreover, it gives information about the believability of the world and the position of the agent within this world.

Thus, we propose to use as a test for an embodied robot a **discrimination task**. Here, the robot has to distinguish between having simulated or 'real' sensory information. This distinction can be easily performed by humans using different mechanisms related to perception and sensation of the world (see (Stadler & Kruse 1986)). We assume that the development of a 'conception' of the body, which is generally discussed as the acquisition of a 'body image' or 'body schema' (see next section), is necessary for embodied action and cognition. The conception of 'body sensation' and awareness can be used to distinguish simulated and real stimuli, e.g. observing the consequences of one's actions, and interactions with the environment, on the body. Additionally, the morphology of a natural body has evolved in phylogeny in close interdependency with, and mutual adaptation to the animate and inanimate environment. In the same way in ontogeny the body is shaped by the environment on the basis of given dispositions (see Maturana's ideas on structural coupling). Transferring this idea to artifacts implies the careful design and adaptation of a robot's body to the environment. In principle the design itself has to be evolvable and self-redesignable. An ideal solution would be bridging of the gap between (robot) hardware and software level, although the idea of realizing evolvable robots with an adaptation of both body and control mechanisms in a close interdependency is still a future goal.

Body Images

A huge amount of literature about body images have been published, but its definition varies widely. According to Slade (Slade 1994) 'body image' should be viewed as a "loose mental representation of the body's shape, form and size". He identified at least seven sets of factors which affect its development and manifestation, including historical, cultural and social, individual and biological factors. Most of these factors are highly dynamic, i.e. varying over time on different time scales. Slade regards the history of sensory input to body experience as the basis for the general 'mental representation of the body'. A combination with at least another kind of variable, the biological variables which might change from day-to-day, indicates the time-changing manifestation of 'body image'. Therefore we suggest a 'dynamical system' point of view of 'body image'. Instead of viewing a 'body image' as a static, structural entity within the brain (a typical 'model' which can, for instance, be implemented in schemata-like representations which are traditionally used in many artificial intelligence approaches) it might be useful to treat 'body image' as a specific manifestation of the dynamics of a state-space built up by cognitive and bodily variables. This point is elaborated further in a following section.

Individual bodies

Every natural system has a unique body and a unique cognition. Even twins with identical genome equipment are not 'equivalent' (in a mathematical sense). Embodied cognition depends on the experiences the individual collects during his/her lifetime. For two individuals there cannot be identical experiences. Even if they are physically close together their viewpoint cannot be identical. This leads us to one of several consequences of being an embodied system: Bodies are physical entities and occupy a distinct volume in space at a specific time. No other individual can be at the same point in space and time. In our perception of human 'reality' we use the metaphor of a 3-dimensional space for locating and interacting with objects. Our body, too, the movements of its limbs and interactions with other objects are located in this geometrical space. The shape of the body changes during active movements or contacts with other objects. The body shows characteristic symmetrical and asymmetrical features concerning movement and orientation of the sensors. Therefore individuals take a certain perspective according to the orientation of the body. In order to preserve the integrity of the body important mechanisms for organisms are to detect contact with or distance to other objects. The active exploration of the environment through body movement is highly important for learning about the environment and the development of cognition. The second-hand knowledge of watching the world from a distance can only complement the first-person experiences of actively interacting with the environment.

Body as a social tool

The human species is the species where the most elaborated forms of 'self-manipulation' of the body can be found. This includes decorating the body, actively manipulating its shape (e.g. through increase or decrease of weight) or using it as a 'device' for social communication: using markers on the body in order to indicate the position in a social hierarchy or using the body as a 'social tool' for threatening or as a 'social stage' to present a certain role or attitude. In (Synnott 1993) Anthony Synnott argues that the human body, its attributes, functions, organs and the senses are not 'given', but socially constructed. The body should be seen as a "social category with different meanings imposed and developed by every age and by different sectors of the population".

The human species is also the species where the most complex social interactions can be found. As a consequence members of a social group must get used to having conspecifics more or less close to their own body. Certain 'regions' (distances around the body) developed in association with the behavior repertoire which can be executed within these regions. The closer conspecifics can approach one's own body, the stronger is usually the degree of familiarization (social bonding) with this person. Therefore social spaces around our body do not only represent the physical distance of an object within. But, in the context of conspecifics, the space is associated with an emotionally colored behavior repertoire. In social interactions the distance to a group member has to be judged according to the social status and the goals and interests of both individuals.

Experiential, bodily understanding

> "Das wissenschaftliche Denken - ein Überblicksdenken, ein Denken des Gegenstands in seiner Allgemeinheit - muß sich in ein vorausgehendes 'Es gibt' zurückversetzen, in die Landschaft, auf den Boden der sinnfälligen Welt und der bearbeiteten Welt, wie sie in unserem Leben, für unseren Körper vorhanden sind, nicht für jeden möglichen Körper, den man, wenn man will, als eine Informationsmaschine betrachten kann, sondern für diesen tatsächlichen Körper, den ich den meinen nenne, diesen Wachtposten, der schweigend hinter meinen Worten und meinen Handlungen steht." ((Merleau-Ponty 1984), cited in (Becker 1997))

In (Galbraith 1995) Mary Galbraith contrasts the different meanings of understanding in computationalism and the continental 'verstehen' Tradition. In the latter, phenomenological approach bodily experiences and social interaction are central to understanding, thought, perception. Experiencing is here described as the "concretely present flow of feelings". According to Galbraith computationalism is based on an explanatory notion of understanding, created by Locke and Kant and with great impact on the development of methodological formalism in natural sciences. Understanding in this way is "detached logical and propositional knowledge of phenomena from the position of a spectator, guaranteed to be valid by experimental repeatability." The notion of understanding in the *verstehen* tradition (influenced by Vico, Schleiermacher, Dilthey, Heidegger, Merleau-Ponty) originated in hermeneutic, historical sciences and arts. In this notion of understanding we understand each other's actions and words not as physically caused, but as a dialogical relationship in which we interpret meaning from each other's gestures on the basis of our reality (continuous, lived experience amid humanly meaningful contexts). The understanding of a novel which leads to involvement, new lived experience and new understanding (in a unique way for each individual at a certain period of time) is an example. Merleau-Ponty's philosophy of mind and language uses the concept of a *lived body*, "We inhabit language because and in the the same way we inhabit our bodies in the world. One enters into language, as one enters into the physical world, by taking up a bodily position within it. Understanding the world and language happens through living it" (see (Becker 1997) for a discussion of Merleau-Ponty's philosophy). Mary Galbraith mentioned Merleau-Ponty and (as an example for a contemporary researcher) the neuropsychologist Oliver Sacks as both arguing to combine (or *envelop* as Merleau-Ponty put it) phenomenology and scientific, empirical studies.

As Barbara Becker (Becker 1997) points out the invention of new methods in artificial intelligence and cognitive science for modelling cognition like connectionism, distributed artificial intelligence, artificial life have not changed the global 'program' of computationalism, namely that computer programs are interpreted as theoretical models or theories (or at least as heuristics) which give *explanations* for human cognitive processes. In general we agree to this viewpoint, but we think that these methods, especially approaches in artificial life which are concerned with the construction of embodied robotic agents, when used in a different context *can* provide a way to overcome the computationalistic/phenomenological gap (a forthcoming paper will elaborate this point is much more detail).

Towards Embodied Cognition

The discussions in this section have been strongly influenced by recent research in dynamical systems.

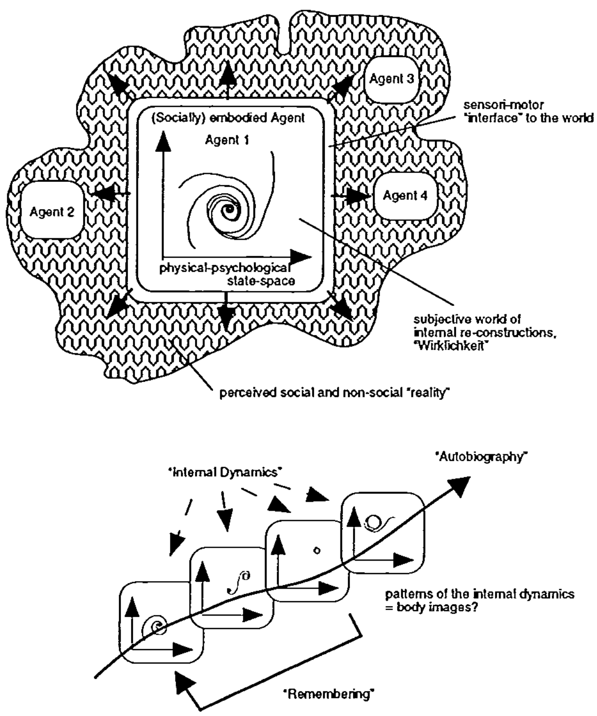

Figure 1: The concept of the physical-psychological state-space

We assume that there are limitations to many dynamic systems approaches towards embodied agents, namely the assumption that the state spaces are built out of dimensions which *represent* parameters of the body. In robotic research sensors are used to measure the physical state of the body (external and internal sensors) and these abstractions are used for calculations within a mathematical framework. This one-way-direction leads to problems which are frequently discussed especially in artificial intelligence, including e.g. debates about the best kind of representation (the propositional versus analog debate) and the 'frame problem'. We think that many problems arise from the 'representation point of view', i.e. the assumption that physical states can be measured, modeled and represented and used for further algorithmic or symbolic calculations without providing a direct way back to the physical basis. This should be possible in a multi-dimensional state-space as a physical-psychological unity which is an interaction-space of agent-environment and of the processes going on within the agent which is situated in a world with dynamic social and non-social interactions (see fig. 1). The main underlying assumptions and characteristics of this concept are summarized as follows:

a) There are no abstract, disembodied models in the brain of humans or other natural systems. Physical states are not transformed to abstract values in order to build up internal models.

b) The physical-psychological state-space consists of dimensions with links to physical states (location of body parts, proprioception, emotional states etc.). The dimensions of the state-space consist of direct physical links to the body, as they are found in humunculi-like projections in the brain.

c) In the physical-psychological state-space trajectories are perceptions, actions, and 'mental units' of thought and remembering. In this way there is no difference between *storage* and *retrieval* of entities. Instead both refer to a *construction* process. In this state-space all experiences are intuitively related to the body and its history.

d) Remembering means *re-construction* of a trajectory, which is (by definition of the state-space) related to the *actual* physical state of the body. The actual physical state is pushed towards a state which results in a trajectory as close as possible to the remembered one. Therefore there are no two identical results of 'memory retrieval'. Every retrieval process means rewriting and modifies its content (see (Rosenfield 1993)). Learning is not seperated from remembering.

e) If bodily characteristics or the physical links between body and state-space change, the dimensions of the state-space change accordingly. In the same way after amputations of body parts, e.g. fingers, the sensorimotor projections in the brain change. A set of dimensions have links to internal emotional states. Changes in these domains might therefore dominate, suppress or activate other domains and lead to crucial changes of the whole state-space. Body images both of oneself and of others could be characterized by attractors in the physical-psychological state-space. This might explain why changes in the emotional states might lead to distortion of the conception of one's body. In the same way certain physical behaviors or habits might influence the emotional attitude towards the body.

A dynamic systems approach has already gained much attention in child development. For instance in (Smith & Thelen 1993) different researchers try to model and analyse existing data from child development in the framework of 'dynamic systems'. The hypothesis that cognitive systems are dynamical systems and can be understood as such (in contrast to the computational hypothesis, namely that cognitive systems are computers) is outlined in detail by Tim van Gelder in (van Gelder 1996). In the case of social understanding one of the challenging points for specifiying a model according to the dynamical systems approach could be to identify an appropriate level and defining variables and parameters (which can provide the coupling to the social environement) appropriately in order to capture the essential aspects of empathic re-experiencing and re-construction.

It is not at all clear how a dynamic model of a 're-experiencing robotic mind' can look like and how it could be implemented. Nevertheless, the metapher of dynamic systems approach has already be used for controlling robots, e.g. when the behavior of robots is modeled as attractors (Steels 1994) and can be used for the development of concepts (Jaeger 1995).

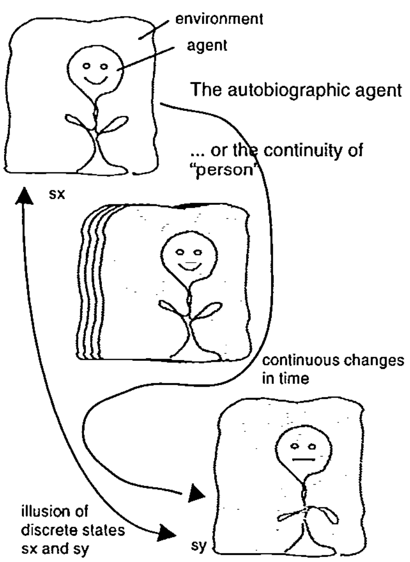

Figure 2: The concept of an autobiographic agent. Dynamically, continiously changing agent-environment interactions give rise to a unique, individual history.

The autobiographic agent

On the basis of the aspects which we described in the previous sections we like to introduce the concept of an *autobiographic agent* (see fig. 2).

To our view the constructive remembering and recollection processes should always been seen in the life-long perspective, i.e. referring to the autobiographic aspect as an ongoing re-construction of the own history and creating the concept of individual personality. The approach to focus on the life-long aspect of human memory is in line with research which has been done in psychology on autobiographic memories (e.g. (Conway 1996a)). The topic memory should not be the central focus of this paper, but nevertheless we see it to be a crucial point in the construction of cognitive systems and modelling of social dynamics (see (Dautenhahn & Christaller 1996) for details). We think that there are still a lot of things left that engineers can learn from remembering in natural systems, which (as it becomes more and more obvious) is a central part for human cognition. The way how impairments of short- or long-term memory part could influence behavior and personality can for instance be found in (Sacks 1985). A constructivist, dynamic account of remembering seems to be most promising and a lot of research is done in this field. To our view this should be linked to the aspect that "memory is primarily a vehicle for personal meanings and for grounding of the self, and that accuracy is secondary to this role" (Conway 1996b). An important aspect in artificial intelligence research on knowledge and memory is *consistency*. Various algorithms have been developed in order to manage a consistent knowledge or data base. On the other hand, humans easily seem to cope with this problem. But there is much evidence that the problem of consistency itself is an artificial one. Instead, the subjective impression of being a static 'personality' is an illusion and might only be a good approximation on small time-scales, (Bartlett 1932). Humans seem to integrate and interpret new experiences on the basis of previous ones. Previous experiences are reconstructed with the actual body and context as the point of reference. In this way past and presence are closely coupled. In combination with human capabilities of rehearsal (as the basis for acting and planning) this coupling is linked to the future. Humans do not seem to worry much about consistency, they give explanations for their behavior on the basis of a story, a dynamically updated script, their autobiography. Believability of this story (to both oneself and others) seem to be more crucial than consistency.

In order to account for this autobiographic aspect of the individual we define the concept of an *autobiographic agent* as an embodied agent which dynamically reconstructs its individual 'history' (autobiography) during its life-time. We find this term useful as a refinement of the general term 'agent' which is only loosely defined and does not help people to identify common research interests. In (Dautenhahn 1996) we discuss in more detail the role of an autobiographic agent in relationships to social systems on different levels of abstractions (the individual on the component level of social systems up to the individuals as part of larger social systems).

Conclusion

The argumentation in this paper can be summarized as follows:

1) Embodiment is linked to a *concept* of a body and is not necessarily given when running a control program on robot hardware.

2) We introduced the concept of 'body image' which is in our conception a dynamically re-constructed pattern (attractor) capturing both robot-environment interactions as well as well as the robots 'internal dynamics'.

3) Embodiment should always be seen as a characteristic of an *individual* and *socially embedded* cognitive system.

4) Studying embodiment for artifacts might contribute to bridge the gap between phenomenological understanding (which has always been grounded on a lived body) and the computationalistic approaches which traditionally dominate computer science and cognitive science. The study of experiential, bodily understanding could be a promising path towards intelligent arti-

facts.
5) We outlined the framework of the *physical-psychological state-space* which changes dynamically in a self-modifying way is a holistic approach towards embodied human and artificial cognition. This is opposed to other structuralist approaches which are well known in artificial intelligence and cognitive science, namely, assuming different cognitive modules located inside a system. The metapher of a *physical-psychological state-space* is meant as a basis for discussion towards a common framework for cognitive architectures comprising natural and artificial systems.
6) In order to stress the importance of a dynamic memory we introduced the concept of an 'autobiographical agent'. It is defined as an embodied agent which dynamically reconstructs its individual 'history' during its life-time. In this way an embodied cognitive agent has to be good at storytelling (telling believable stories of its own autobiography).

Acknowledgements

I like to thank Barbara Becker for valuable hints and comments on philosophical issues, especially on the embodiment of human cognition. I am grateful to Thomas Christaller for discussing with me constructively the concept of a physical-psychological state-space. The research and writing of this paper has been financed by a HCM/TMR research grant from the European Community.

References

Bartlett, F. C. 1932. *Remembering – A study in experimental and social psychology.* Cambridge University Press.

Becker, B. 1997. Leiblichkeit und Kognition. Anmerkungen zum Programm der Kognitionswissenschaften. In Engel, A., and Gold, P., eds., *Der Mensch in den Kognitionswissenschaften.* Suhrkamp.

Brooks, R. A., and Mataric, M. J. 1994. Real robots, real learning problems. In H., C. J., and S., M., eds., *Robot learning.* Kluwer Academic Publishers. 193–213.

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

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

Conway, M. A. 1996a. Autobiographical knowledge and autobiographical memories. In Rubin, D. C., ed., *Remembering out past. Studies in autobiographical memory.* Cambridge University Press. 67–93.

Conway, M. A. 1996b. Shifting sands. *Nature* 380:214.

Dautenhahn, K., and Christaller, T. 1996. Remembering, rehearsal and empathy - towards a social and embodied cognitive psychology for artifacts. In O'Nuallain, S., and Kevitt, P. M., eds., *Two sciences of the mind. Readings in cognitive science and consciousness.* John Benjamins North America Inc. 257–282.

Dautenhahn, K. 1995. Getting to know each other – artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems* 16:333–356.

Dautenhahn, K. 1996. Social modelling – on selfish genes, social robots and the role of the individual. submitted to Int Journal in Computer Simulation, special issue on the Simulation of Social Behavior.

Galbraith, M. 1995. The *verstehen* tradition. *Minds and Machines* 5:525–531.

Jaeger, H. 1995. Identification of behaviors in an agent's phase space. Technical report, series Arbeitspapiere der GMD.

Merleau-Ponty, M. 1984. *Das Auge und der Geist.*

Pfeifer, R. 1995. Cognition – perspectives from autonomous agents. *Robotics and Autonomous Systems* 15:25–46.

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

Roth, G. 1994. *Das Gehirn und seine Wirklichkeit.* Suhrkamp.

Sacks, O. 1985. *The man who mistook his wife for a hat.* Summit Books.

Slade, P. D. 1994. What is body image? *Behaviour Research and Therapy* 32(5):497–502.

Smith, L. B., and Thelen, E., eds. 1993. *A dynamic systems approach to development: applications.* A Bradford Book, The MIT Press.

Stadler, M., and Kruse, P. 1986. Gestalttheorie und Theorie der Selbstorganisation. *Gestalt Theory* 8:75–98.

Steels, L. 1994. Mathematical analysis of behavior systems. In Gaussier, P., and Nicoud, J.-D., eds., *Proc. From Perception to Action Conference, Lausanne, Switzerland,* 88–95. IEEE Computer Society Press.

Steels, L. 1996. The spontaneous self-organization of an adaptive language. S. Muggleton (ed.): Machine Intelligence 15, Oxford University Press.

Synnott, A., ed. 1993. *The body social.* Routledge.

van Gelder, T. 1996. The dynamical hypothesis in cognitive science. to appear in Behavioral and Brain Sciences.