Ziemke, T. (1996b) Towards Autonomous Robot Control via Self-Adapting Recurrent Networks. *Artificial Neural Networks - ICANN 96*, pp. 611-616. Berlin/ Heidelberg, Germany: Springer Verlag.

Ziemke, T. (1996c) Towards Adaptive Perception in Autonomous Robots using Second-Order Recurrent Networks. *Proceedings of the First Euromicro Workshop on Advanced Mobile Robots (EUROBOT '96)*, pp. 89-98. Los Alamitos, CA: IEEE Computer Society Press.

Ziemke, T. (1997) The 'Environmental Puppeteer' Revisited: A Connectionist Perspective on 'Autonomy'. *Proceedings of the 6th European Workshop on Learning Robots (EWLR-6)*. Brighton, UK, August 1997.

Ziemke, T. (1998) Adaptive Behavior in Autonomous Agents. *Presence*, 5(6).

Ziemke, T. & Sharkey, N. E. (eds.) (1998) *Biorobotics*. Special issue of *Connection Science*, 10(3-4).

University of Texas at Austin.

Loeb, J. (1918) *Forced movements, tropisms, and animal conduct*. Philadelphia: Lippincott Company.

Lund, H. H.; Hallam, J. & Lee, W. (1997) Evolving robot morphology. *Proceedings of the IEEE Fourth International Conference on Evolutionary Computation*. IEEE Press.

Maturana, H. and Varela, F. (1987) *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, MA: Shambhala.

Mondada, F. & Floreano, D. (1995) Evolution of neural control structures: Some experiments on mobile robots. *Robotics and Autonomous Systems*, 16(2-4): 183-196.

Newell, A. (1980) Physical Symbol Systems. *Cognitive Science*, 4: 135-183.

Newell, A. & Simon, H. (1976) Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19: 113-126.

Nolfi, S. (1997a) Using emergent modularity to develop control systems for mobile robots. *Adaptive Behavior*, 5(3-4): 343-363.

Nolfi, S. (1997b) Evolving Non-Trivial Behavior on Autonomous Robots: Adaptation is More Powerful Than Decomposition and Integration. In Gomi, T. (ed.), *Evolutionary Robotics - From Intelligent Robots to Artificial Life*. Kanata, Canada: AAI Books.

Peschl, M. F. (1996) The Representational Relation Between Environmental Structures and Neural Systems: Autonomy and Environmental Dependency in Neural Knowledge Representation. *Nonlinear Dynamics, Psychology and Life Sciences*, 1(3).

Pfeifer, R. (1995) Cognition - Perspectives from autonomous agents. *Robotics and Autonomous Systems*, 15: 47-70.

Pfeifer, R. & Scheier, C. (1998) *Understanding Intelligence*. Cambridge, MA: MIT Press.

Port, R. & van Gelder, T. (1995) *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Regier, T. (1992) *The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization*. PhD Thesis / Tech. Rep. TR-92-062. Berkeley: Dept. of Computer Science, University of California at Berkeley.

Roeder, K. & Treat, A. (1957) Ultrasonic reception by the tympanic organs of noctoid moths. *Journal of Experimental Zoology*, 134:127-158.

Rutkowska, J. C. (1996) Reassessing Piaget's Theory of Sensorimotor Intelligence: A View from Cognitive Science. In: Bremner, J. G. (ed.) *Infant Development: Recent Advances*. Hillsdale, NJ: Lawrence Erlbaum.

Rylatt, M. & Czarnecki, C. (1998) Beyond Physical Grounding and Naive Time: Investigations into Short-Term Memory. In Pfeifer, R., Blumberg, B., Meyer, J.-A. & Wilson, S. W (eds.) *From Animals to Animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pp. 22-31. Cambridge, MA: MIT Press.

Schank, R. C. & Abelson, R. P. (1977) *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum.

Searle, J. (1980) Minds, brains and programs. *Behavioral and Brain Sciences*, 3: 417-457.

Sharkey, N. E. (1997) Neural networks for coordination and control: The portability of experiential representations. *Robotics and Autonomous Systems*, 22(3-4): 345-359.

Sharkey, N. E. & Heemskerk, J. H. (1997) The Neural Mind and the Robot. In A. J. Browne (Ed.) *Neural Perspectives on Cognition and Adaptive Robotics*. Bristol, UK: IOP Press.

Sharkey, N. E. & Jackson, S. A. (1994) Three Horns of the Representational Trilemma. In: Honavar, V. & Uhr, L. (eds.) *Symbol Processing and Connectionist Models for Artificial Intelligence and Cognitive Modeling: Steps towards Integration*, pp. 155-189. Academic Press.

Sharkey, N. E. & Jackson, S. A. (1996) Grounding Computational Engines. *Artificial Intelligence Review*,10: 65-82.

Sharkey, N. E. & Ziemke, T. (1998) A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science*, 10(3-4): 361-391.

Sherrington, C. S. (1906) *The integrative action of the nervous system.* New York: C. Scribner's Sons.

Tani, J. (1996) Does Dynamics Solve the Symbol Grounding Problem of Robots? An Experiment in Navigation Learning. *Learning in Robots and Animals - Working Notes*. AISB'96 workshop, Brighton, UK.

van Gelder, T. J. (1995) What might cognition be if not computation? *Journal of Philosophy*, 91: 345-381.

van Gelder, T.J. (1998) The Dynamical Hypothesis in Cognitive Science. *Behavioral and Brain Sciences*.

Varela, F., Thompson, E. & Rosch, E. (1991) *The Embodied Mind - Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

von Uexküll, J. (1928) *Theoretische Biologie*. Frankfurt/ Main, Germany: Suhrkamp Verlag.

Wilson, S. W. (1991) The Animat Path to AI. *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.

Ziemke, T. (1996a) Towards Adaptive Behaviour System Integration using Connectionist Infinite State Automata. In: Maes, P., Mataric, M., Meyer, J.-A., Pollack J. & Wilson, S. (eds.) *From Animals to Animats 4 - Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pp. 145-154. Cambridge, MA: MIT Press.

Simulation of Adaptive Behavior, pp. 421-429. Cambridge, MA: MIT Press.

Bickhard, M. & Terveen, L. (1995) Foundational Issues in Artificial Intelligence and Cognitive Science - Impasse and Solution. New York: Elsevier.

Biro, Z. & Ziemke, T. (1998) Evolution of visually-guided approach behaviour in recurrent artificial neural network robot controllers. In Pfeifer, R., Blumberg, B. Meyer, J.-A. & Wilson, S. W (eds.) *From Animals to Animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pp. 73-76. Cambridge, MA: MIT Press.

Brooks, R. A. (1986) A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation*, 2: 14-23.

Brooks, R. A. (1989) A Robot that Walks: Emergent Behavior from a Carefully Evolved Network. *Neural Computation*, 1(2): 253-262.

Brooks, R. A. (1990) Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6(1-2): 1-16.

Brooks, R. A. (1991a) Intelligence Without Representation. *Artificial Intelligence*, 47: 139-160.

Brooks, R. A. (1991b) Intelligence Without Reason. *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pp. 569-595. San Mateo, CA: Morgan Kauffmann.

Brooks, R. A. (1993) The Engineering of Physical Grounding. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pp. 153-154. Hillsdale, NJ: Lawrence Erlbaum.

Chalmers, D.J. (1992) Subsymbolic computation and the Chinese room. In: Dinsmore, J. (ed.) *The Symbolic and Connectionist Paradigms: Closing the Gap*. Hillsdale, NJ: Lawrence Erlbaum.

Chiel, H. J. & Beer, R. A. (1997) The brain has a body: Adaptive Behavior emerges from interactions of nervous system, body, and environment. *Trends in Neurosciences*, 20:553-557.

Clark, A. (1997) *Being There - Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. & Wheeler, M. (1998) Bringing Representation Back to Life. In Pfeifer, R., Blumberg, B. Meyer, J.-A. & Wilson, S. W (eds.) *From Animals to Animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pp. 3-12. Cambridge, MA: MIT Press.

Cliff, D. & Miller, G. F. (1996) Co-evolution of Pursuit and Evasion II: Simulation Methods and Results. In: Maes, P., Mataric, M., Meyer, J.-A., Pollack J. & Wilson, S. (eds.) *From Animals to Animats 4 - Proceedings of the Fourth International. Conference on Simulation of Adaptive Behavior*, pp. 506-515. Cambridge, MA: MIT Press.

Cottrell, G. W., Bartell, B. & Haupt, C. (1990) Grounding Meaning in Perception. *Proceedings of the German Workshop on Artificial Intelligence (GWAI)*, pp. 307-321.

Dorffner, G. & Prem, E. (1993) Connectionism, Symbol Grounding, and Autonomous Agents. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pp. 144-148. Hillsdale, NJ: Lawrence Erlbaum.

Floreano, D. (1997) Reducing Human Design and Increasing Adaptability in Evolutionary Robotics. In Gomi, T. (ed.), *Evolutionary Robotics - From Intelligent Robots to Artificial Life*. Kanata, Canada: AAI Books.

Fodor, J. A. (1980) Methodological solipsism considered as a research strategy in cognitive science. *Behavioral and Brain Sciences*, 3: 63-110.

Fodor, J.A. (1981) *Representations: philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.

Fodor, J. A. (1983) *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3-71.

Funes, P. & Pollack, J. B. (1997) Computer evolution of buildable objects. In: Husbands, P. & Harvey, I. (eds.) *Proceedings of the Fourth European Conference on Artificial Life*, pp. 358-367. Cambridge, MA: MIT Press.

Globus, G. G. (1992) Toward a Noncomputational Cognitive Neuroscience. *Journal of Cognitive Neuroscience*, 4(4).

Gruau, F. (1995) Automatic definition of modular neural networks. *Adaptive Behavior*, 2: 151-183.

Harnad, S. (1989) Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence*, 1: 5-25.

Harnad, S. (1990) The Symbol Grounding Problem. *Physica D*, 42: 335-346.

Harnad, S. (1993) Symbol Grounding is an Empirical Problem: Neural Nets are Just a Candidate Component. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pp. 169-174. Hillsdale, NJ: Lawrence Erlbaum.

Harnad, S. (1995) Grounding Symbolic Capacity in Robotic Capacity. In: Steels, L. & Brooks, R. A. (eds.) *The "artificial life" route to "artificial intelligence". Building Situated Embodied Agents*, pp. 276-286. New Haven: Lawrence Erlbaum.

Lakoff, G. (1993) Grounded Concepts Without Symbols. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pp. 161-164. Hillsdale, NJ: Lawrence Erlbaum.

Law, D. and Miikkulainen, R. (1994) *Grounding Robotic Control with Genetic Neural Networks*. Tech. Rep. AI94-223. Austin: Dept. of Computer Sciences, The

spective approaches to grounding, however, despite their differences, a number of points of 'convergence' can be noted: Both categories of grounding approaches require their agents to have *robotic capacities*; in the cognitivist framework these are somewhat peripheral but necessary; in the enactive framework robotic capacities are at the core of the view of cognition as embodied action. Furthermore, as we have argued here, both approaches require truly grounded systems to be '*complete agents*': In the cognitivist approach grounding requires input and central systems embedded in their environment. In the enactive framework full grounding or rooting requires agents to have developed as a whole in interaction with their environment. Finally, both types of approaches rely on a certain degree of *bottom-up development/evolution*: In the cognitivist approach both development and evolution are required to account for grounding of both innate and learned representations, input systems, etc. In the enactive framework radical bottom-up development, at both individual and species level, of integrated embodied agents seems essential to creating artefacts with the rooting and environmental embedding that forms the basis of intelligent behaviour and cognition in living systems.

For the enactive approach to AI, there are a few more practical lessons to be drawn from the discussion presented in this paper. Firstly, it has been argued here that the enactive/robotic AI research community will have to do some rethinking of its 'cornerstones':

- *Natural embodiment* is more than being-physical. In addition to that it reflects/embodies the history of structural coupling and mutual specification between agent and environment in the course of which the body has been constructed.
- *Natural situatedness* is more than being physically connected to your environment. It also comprises being embedded conceptually in your own phenomenal world (*Umwelt*), which is also constructed in the course of the above history of interaction, in congruence with sensorimotor capacities as well as physiological and psychological needs

Secondly, despite its commitment to embodied agents the enactive approach is not at all immune to the grounding problem. In fact the opposite it true: Because it recognizes the embodied nature of intelligent behaviour, the enactive approach to AI faces an even harder grounding problem than its traditional counterpart. In cognitivist AI the relation between agent and environment is at least rather well defined, namely representation (in the traditional sense). The cognitivist grounding problem is therefore reduced to a somewhat technical problem, namely hooking individual objects in external reality to their internal representations. In enactive AI research, however, there just is no such clear 'interface' between agent and environment. As discussed above (and more detailed in Sharkey & Ziemke (1998)), the complex and intertwined relation between natural agents and environments is rooted in a history of structural coupling, and the two mutually influence each other in a multitude of ways. The conceptual core problem for enactive AI therefore is the question of how, if at all, we could build, or rather enable self-organization of, agents that are equally embedded and rooted in their environments.

Clark (1997) recently illustrated his notion of embodied, active cognition with a quote from Woody Allen: "Ninety percent of life is just being there." The arguments presented in this paper could be summarized by saying that the problem with modern AI is that its robots, although physically grounded, still lack the rooting that allows living organisms to just *be there*. Thus the key problem in the attempt to create truly grounded and rooted AI systems is first and foremost the problem of '*getting there*', i.e. the question how, if at all, artificial agents could construct and self-organize themselves and their own environmental embedding.

Acknowledgements

The author would like to thank Noel Sharkey and Zoltán Biró for helpful comments on earlier versions of this paper.

References

Agre, P. E. & Chapman (1987) Pengi: An Implementation of a Theory of Activity. *Proceedings of AAAI-87*. Menlo Park, CA: AAAI, pp. 268-272.

Beer, R.D. (1995) A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72: 173-215.

Beer, R. A. (1996) Toward the Evolution of Dynamical Neural Networks for Minimally Cognitive Behaviour. In: Maes, P., Mataric, M., Meyer, J.-A., Pollack J. & Wilson, S. (eds.) *From Animals to Animats 4 - Proceedings of the Fourth International Conference on*

tation as discussed in this paper, seems to suggest, that in fact we have to be very careful about such abstractions when studying/modelling intelligent behaviour in artefacts, since any abstraction imposes extrinsic (which, however, does not necessarily equal 'wrong') design constraints on the artefact in question, and we will have to re-examine some of the 'details' which perhaps prematurely have been abstracted from earlier.

One of these 'details' is, as mentioned above, the role of the living body. Embodied AI has (rightly) acknowledged the role of the physical body and its causal connection to the environment. It has, however, so far largely treated the body as some sort of physical interface between controller and environment, but ignored the special role the living body plays in the interaction between organism and their environment.

Sharkey & Ziemke (1998) discuss in detail the relation between the work of Sherrington (1906), Loeb (1918) and von Uexküll (1928) and recent work in embodied AI and cognitive science. The key points for the argument at hand are: (1) Living organisms are highly integrated and coherent systems, i.e. different parts of organisms interact in solidarity in a way that allows the whole to act as an integrated individual. (2) By means of its body an organism is embedded not only in a physical environment, but, more importantly, in its own *Umwelt* (von Uexküll 1928) or phenomenal world or "effective environment" (Clark 1997), namely a subjective abstraction, interpretation or constructed view of the physical environment that fits the agent's sensorimotor capacities and its physiological and psychological needs. That means, organisms are 'tailor-made' to perceive and act in intrinsically meaningful ways. A simple example of this are noctoid moths, which have specially tuned 'ears', which, for example, when faced with loud high frequency emissions of nearby bats trigger a desynchronization of wingbeats and thus lead to unpredictable escape behaviour (Roeder & Treat 1957; cf. also Sharkey & Ziemke 1998). (3) Due to the two-way fit between living bodies and their environment the two form a systematic whole, which must be considered the basis of intelligent behaviour and meaningful interaction between them.

Obviously the above aspects of organisms and their embodiment are lacking from the typical AI robot which is rather arbitrarily equipped with ultrasonic and infrared sensors all around its body, because its designers or buyers considered that useful (i.e. a judgement entirely extrinsic to the robot, and grounded, at most, in human design, possibly including *human* experience from, for example, earlier experiments). Despite the emphasis on embodiment in the enaction paradigm and despite the biological inspiration/motivation behind much of modern robotics (see Ziemke & Sharkey (1998) for a number of examples), this type of historical rooting and environmental embedding through a living body, as a result of co-evolution/-development and mutual determination of body, nervous system and environment, has been largely neglected so far in embodied AI research.

A small number of researchers have, however, begun to study the evolution of physical structures and robot morphologies (e.g. Funes & Pollack 1997; Lund et al. 1997), in some cases in co-evolution with controllers, as, for example, the work of Cliff & Miller (1996), in which co-evolution of 'eyes' (optical sensors) and 'brains' (control networks) has been applied (in simulation) to pursuing and evading agents. The approach of (co-) evolutionary robotics is still very young, and its potential and limitations are by far not fully explored yet. This line of research might, however, be a first step to developing robotic agents with (some of) the integration and coherence of living organisms, by rooting them in their environments through co-evolution of robot bodies, control systems and their environments.

Summary and Conclusion

Both cognitivist and enactive approaches to grounding, although in different ways, to some extent follow Searle's conclusion that intelligence is a property of machines, i.e. embodied systems, causally connected with their environment, rather than disembodied computer programs. The enactive approach certainly follows this route more wholeheartedly, with embodiment and agent-environment interaction being at the core of the enactive view of cognition. In the cognitivist approach on the other hand, grounding is rather considered to supply the necessary interface between the external physical world and the internal cognitive processes, which, if cognitivist grounding worked, could still be purely computational.

The question whether cognitivism or enaction is 'right' is beyond the scope of this paper. In their re-

comings. Hence, let us briefly recapitulate the major points so far.

Summary So Far

Searle's (1980) and Harnad's (1990) analyses of work in traditional, purely computational AI showed that programming knowledge into a system alone can never make a system intelligent, since the knowledge will always remain extrinsic to the system, i.e. it will only be actual 'knowledge' to external observers but lack what Rylatt & Czarnecki (1998) call "contents-for-the-machine". Hence, a natural conclusion is that knowledge must only enter a system from its environment in a grounded fashion.

In the cognitivist framework, where (a) 'knowledge' by definition consists of explicit, manipulable, internal representations, and (b) a distinction is made between perceptual input systems (transducing sensory percepts onto internal representations) and central systems (manipulating internal representations), this means (cf. Harnad's (1990) proposal) that any new internal representation must be

- either definable by sensory or sensorimotor invariants (in the case of atomic representations)
- or constructible from already existing atomic or complex representations (in the case of complex representations).

Typically cognitivist grounding approaches, here exemplified with Regier's (1992) work, therefore count on transducing sensory percepts, typically through connectionist networks, onto categorical representations which can then have a 1:1 relation to internal symbolic representations (cf. also Harnad 1990). Problems typically ignored in this approach are that

- the transducing input system, since alone it cannot provide grounding to more than the result of the transduction, has to be embedded in its usage through central systems which themselves have to be embedded in an environment, and
- it cannot be denied that in Regier's system a lot of *his* knowledge went into the design of his transducer (a structured connectionist net), which therefore (according to the above line of reasoning) has to be said to be extrinsic to the overall system.

In the enactive framework, where the agent as whole must be considered to embody 'knowledge', it is more difficult to pin down what exactly has to be grounded. Some degree of physical grounding can be said to come with the sensorimotor embedding of robotic agents in their environment. Further grounding of (effective) behaviour is achieved by adequately transducing sensory percepts onto motor output. Here (transformation) knowledge needs to be embodied in the transducing agent function in order to ensure adequate action: In Brooks's subsumption architecture this knowledge is designed/programmed into the system (resulting in the disadvantages discussed above), whereas using connectionist networks or evolutionary algorithms it can partly be self-organized in a grounded fashion, i.e. acquired in interaction with the environment. We have, however, argued briefly that, due to the fact that robot bodies, unlike living bodies, typically are the results of (external) design rather than self-organization, conventional robots lack the historical *rooting* and embedding that form the basis of intelligent behaviour and meaningful interaction between living organisms and their environments.

From Grounding to Rooting

If we aim for artefacts that are grounded/rooted/embedded in their environments in the sense organisms are, i.e. systems whose behaviour and underlying mechanisms are in fact intrinsic and meaningful to themselves, then we have to go beyond grounding designed artefacts by 'hooking' them to pre-given environments, and have to start looking at systems which as a whole have developed in interaction with their environment, and thus are truly rooted in it.

In fact, the only truly intelligent systems we know of are animals, i.e. biological systems whose genotype has evolved over millions of years, and who in many cases undergo years of individual development before achieving full intelligence. Thus, animals are embedded in their environments in a multitude of ways, whereas most grounding approaches rather aim for hooking pre-given agents to pre-given environments, by means of representations or effective behaviour.

AI and cognitive science, in their attempt to synthesize and model intelligent behaviour, have always been based on high-level abstractions from the biological originals (disembodiment, the 'information processing metaphor', the 'brain metaphor', etc.). The grounding problem, in its broad interpre-

of units, layers, etc. in connectionist networks) the designer will necessarily impose structural constraints on the system, in particular when designing modular or structured control mechanisms (cf. Nolfi 1997a, 1997b; Ziemke 1996b).

Grounding Control Structures: An approach to further reduce determination through human design, is to ground not only internal parameters of control mechanisms but also their structure in agent-environment interaction, e.g. through evolution of connectionist control architectures (e.g. Floreano 1997; Mondada & Floreano 1994; Gruau 1995). One such approach to ensure grounding of robotic control while limiting restrictions imposed through design to a minimum is the work by Law & Miikkulainen (1994), who let connectionist architectures (to be exact: the connectivity in a given architecture) evolve, thereby grounding the actual network architecture in experience (to some extent). Law and Miikkulainen argued that

> ... the agents that are the product of this system will be undeniably grounded in their simulated world, since they will have begun from ground zero, knowing nothing at all.[2] (Law & Miikkulainen 1994, footnote added)

Another approach that partly addresses the problem of grounding control structure is the author's work on 'self-adapting' recurrent connectionist robot controllers (Ziemke 1996a, 1996c) in which the sensorimotor mapping is actively (re-) constructed in every time step by a second connectionist net. This enables the overall controller to exhibit an emergent, grounded task decomposition (cf. also Nolfi 1997b) and autonomously acquire a corresponding self-organized virtual modularisation. This allows the controlled robot to exhibit different behaviours at different points in time, without these behaviours or their relation and organization actually being built into the system. A similar approach, although using a different network architecture, was used by Biro & Ziemke (1998) who evolved recurrent connectionist control networks to exhibit subsumption-architecture-like organization of different behaviours without such structure actually being built into the control mechanisms.

A problem with grounding control systems, or even their structure, in experience and agent-environment interaction is what Funes & Pollack (1997) called the "chicken and egg" problem of adaptive robotics:

> Learning to control a body is dominated by inductive biases specific to its sensors and effectors, while building a body which is controllable is conditioned on the pre-existence of a brain. (Funes & Pollack 1997)

In other words, for example, the weights in a trained connectionist robot controller could be considered grounded; they are, however, meaningful only in the context of the robot body, sensors, motors, etc. and their embedding in the environment (cf. Sharkey & Ziemke 1998). The body, and thus the agent's environmental embedding, however, are in the vast majority of cases in robotic AI still 'provided' to the agent by an external designer, and therefore, following the above arguments, have to be considered extrinsic to the agent itself.

Thus enactive AI research is facing its very own variation of the grounding problem, namely what might be called the *robot grounding* or *body grounding problem*. We have argued elsewhere (Sharkey & Ziemke 1998) in detail that the Brooksian notions of embodiment and physical grounding discussed above, which belong to the foundations of enactive AI and modern robotics, fail to fully capture the way living systems, by means of their bodily mechanisms, are embedded in their environment. Organisms and their environments are not designed separately and then "hooked" together. A living body provides much more than physical grounding, and, unlike a conventional robot body, a living body embodies a long history of mutual specification and structural coupling of organism and environment in the course of evolution and the individual organism's lifetime. Thus any organism is deeply historically *rooted* in its environment, and the two form a meaningful whole, which is the basis of the delicate and complex interplay exhibited by living systems and their environments. This point will be elaborated further in the following sections, but for a detailed discussion of this aspect see also Sharkey & Ziemke (1998).

Discussion

This paper has so far given a 'guided tour' around the grounding problem and a number of approaches aimed at solving it, all of which, however, at least in the author's opinion, have their problems and short-

[2] Note however that sensors, motors, and some knowledge of their availability are still built into the system.

tically speaking, however, this approach to constructing the agent function could as well be characterized as incremental trial-and-error engineering, bringing with it, no matter how carefully it is carried out, the limitations of designing/engineering which we already noted in the discussion of Regier's work: The system's actions could be considered grounded in its environment (which causally participates in producing the behaviour), the internal mechanisms realizing the agent function (that is, the behavioural modules and their interconnection), however, are in no way intrinsic to the system.

The same problem was noticed earlier in Regier's case, the consequences are, however, more 'dramatic' here: The ungrounded transducer in Regier's case was an input system of (arguably) peripheral relevance to the central computational engine, whereas here the ungrounded 'transducer' is the complete agent (function) itself. Hence, the problem here is analog to that in the case of the Chinese Room (as well as that of the pocket calculator): The system might exhibit the 'right' behaviour, its internal mechanisms (its modularisation and the resulting task decomposition, the FSA, etc.), however, are not intrinsic to the system, but are 'only' grounded in careful engineering by an external designer.

Grounding Agent Functions: Physical grounding offers a way for AI research to escape the internalist trap. It does, however, also offer a way into what might be called the *externalist trap*: If it is only the "here and now of the world" (see the above Brooksian notion of situatedness) that determines an agent's behaviour, i.e. if the agent is merely reacting to its current environment, then the agent is best described as controlled by the "environmental puppeteer" (Sharkey & Heemskerk 1997) rather than as an *autonomous* agent (cf. Ziemke 1997; Ziemke 1998). This is also reflected in Pfeifer's (1995) more encompassing definition of a situated agent:

> ... a situated agent is one which can bring to bear its own experience onto a particular situation, and the interaction of its experience with the current situation will determine the agent's actions. ... Note that a situated agent is different from a reactive one. A reactive agent does not incorporate experience - it will always react the same way in the same situation. (Pfeifer 1995)

There are, at least, two ways for an agent to bring to bear experience in determining its own behaviour. Firstly, the agent can 'free' itself (partly) from the 'environmental puppeteer', i.e. dependence on the "here and now of the world", by using an *internal state or memory* in addition to current input, instead of merely reacting to the latter. Rylatt & Czarnecki (1998) point out that physical grounding alone does not account for intrinsic meaning or, as they put it, "contents-for-the-machine". In addition, they argue, agents need to be "embedded in time" through the use of memory. Secondly, an agent can free itself (partly) from its pre-programming by *learning*, i.e. utilize its experience in order to adapt the mechanisms underlying its behaviour in a self-organizing fashion, and thus to further ground its behaviour in agent-environment interaction (e.g. Law & Miikkulainen 1994; Beer 1996). For a more detailed discussion of these two aspects as essential elements of (artificial) autonomy see Ziemke (1998).

Approaches to grounding behaviour in experience therefore typically aim to reduce as much as possible the role of the designer/engineer/programmer in determining how to realize the agent function. The typical approach to grounding an agent function in experience is to connect sensors and actuators through some control mechanism (e.g. a connectionist network or a classifier system) and to let agents adapt the control mechanism on the basis of experience in the course of evolutionary or self-learning. The approach has some obvious advantages, the agent function can now be *self-organized* by the agent, through adjustment of internal parameters (connection weights, classifier strengths, etc.) instead of having to be programmed by an external designer. Hence, the internal parameters of the control mechanism and the resulting behaviour of such a self-organized agent could be considered grounded in experience (e.g. Tani 1996, Beer 1996).

Pfeifer (1995), for example, describes a robot after neural network learning as follows:

> The agent's categorization of the environment, i.e. its prototypes, are *grounded* since they are acquired through its interaction with the environment and are therefore built up from its own point of view, not from one of the observer. (Pfeifer 1995)

The problem of design, however, remains to some degree even in self-organizing control systems, since by choice of architecture (e.g., number

gence, situatedness and embodiment" (Brooks 1991b).

The commitment to machines, i.e. robotic agents, rather than computer programs, as the object of study is reflected in the notion of embodiment:

> **[Embodiment]** The robots have bodies and experience the world directly - their actions are part of a dynamic with the world and have immediate feedback on their own sensations (Brooks 1991b, original emphasis)

These robotic agents are typically considered *physically grounded* (Brooks 1990). That means, they are causally connected to their environment through the use of sensory input and motor output ("immediately grounded representations" according to Dorffner & Prem (1993)) such that, as Brooks (1993) argues, internally "everything is grounded in primitive sensor motor patterns of activation".

The commitment to the study of agent-environment interaction, rather than abstract reasoning and world modelling, is further reflected in the notion of situatedness:

> **[Situatedness]** The robots are situated in the world - they do not deal with abstract descriptions, but with the here and now of the world directly influencing the behavior of the system. (Brooks 1991b, original emphasis)

Physical grounding and agent-environment interaction obviously enable an agent to 'reach out' into its environment and directly interact with it, i.e. they offer a way to escape the internalist trap. Physical grounding does, however, only offer a pathway for hooking an agent to its environment. It does, by itself, not ground behaviour or internal mechanisms (cf. Sharkey & Ziemke 1998; Rylatt & Czarnecki 1998; cf. also Searle's (1980) discussion of the 'robot reply' to the CRA), as will be discussed in detail in the following.

Grounding Behaviour: Instead of the central modelling and control typical for the cognitivist paradigm, enactive systems typically consist of a number of behavioural subsystems or components working in parallel from whose interaction the overall behaviour of a system *emerges*. Accordingly, it is not representations in the traditional sense, but rather an agent's behaviour that has to be grounded in its environment (e.g., Law & Miikkulainen 1994; Beer 1996). (Note however, that, if 'behaviour-generating patterns' are considered representations, then, of course, behaviour grounding also amounts to representation grounding, although of a different type.)

The lack of manipulable world models and representations in the traditional sense in enactive system might at first appear to simplify grounding, since it is exactly this representational 'knowledge' that requires grounding in the cognitivist framework. This does, however, also pose a serious problem, since 'knowledge' in the enactive paradigm, rather than in explicit world models, is typically considered to be embodied in a distributed fashion (body, sensors, actuators, nervous/control system, etc.) or partly even lie in the environment (e.g., Maturana & Varela 1987; Varela et al. 1991; Brooks 1991b; Clark 1997; Chiel & Beer 1997). If an agent's behaviour requires grounding, then obviously the 'behaviour-generating patterns' it results from do so too. The list of elements, however, that participate in generating behaviour basically contains all mechanisms which, in one way or another, participate in the flow of activation from sensors to actuators. Hence, the question here is where to start grounding and where to end it?

Most commonly the grounding of behaviour is approached as a matter of finding the right *agent function*, i.e. a mapping from sensory input (history) to motor outputs that allows an agent to effectively cope with its environment. There are basically two different ways of achieving this, which will be discussed in the following: (a) engineering/designing and the agent function, and (b) self-organizing the agent function, and thus grounding itself in experience.

Engineering Agent Functions: A classical example for the engineering of agent functions is Brooks' (1986) subsumption architecture, in which the overall control emerges from the interaction of a number of hierarchically organized behaviour-producing modules. For example, the control of a simple robot that wanders around avoiding obstacles could emerge from one module making the robot go forward and a second module which, any time the robot encounters an obstacle, overrides the first module and makes the robot turn instead.

In Brooks' own work (see Brooks (1989) for a detailed example) typically each of the behavioural modules is implemented as a finite-state-automaton (FSA), and behavioural competences are carefully and incrementally layered bottom-up in a process which is supposed to mimic, to some degree, the evolution of biological organisms. Less euphemis-

the system, the behaviour has to be *grounded in agent-environment interaction*, just as it was argued earlier (following Harnad) representations had to be. Accordingly, for the above labelling act to make sense to an agent, that agent would have to be able to at least use its spatial labels in some way (e.g., to communicate them to other agents), to profit in some way from developing the capacity to do so, etc.

Cognitivists could of course rightly argue that the functional value of the transduction/labelling act, and thereby its meaning to the overall system, lies in its support of hypothetical central computational systems which could make use of the resulting representation of the labelled object/scene. In Regier's system, however, as discussed above, there just is no such overall system to which the labelling could be intrinsic.

Secondly, assuming there were such central systems, that made the act of transduction intrinsically meaningful to the overall system (consisting of central systems and transducing input system), could we then speak of a truly grounded system? No, we still could not, since *the transducer* (Regier's labelling system) *itself* (its structure, organization, internal mechanisms, etc., basically all of it except for the networks' connection weights) is not grounded in anything but Regier's design ideas, however good and psychologically or neurobiologically plausible they might be.

In this particular case the transducing labelling system is a structured connectionist model using two topographic maps dedicated to processing input for the two objects, and a number of further layers/networks to process the output of these maps. Regier (1992) himself argued that his system is a pre-adapted structured device that basically finds itself confronted with a task similar to that an infant is facing when acquiring lexical semantics for spatial terms. There is, however, at least one major difference, and that is the fact that the corresponding subsystem in humans (to the extent that it is innate) has been pre-adapted, i.e developed and tested to work together with the rest of the human being in an integrated fashion, during the course of evolution, such that it very well could be said to be intrinsic to the human species (or genotype), and thereby to the individual (or phenotype) as an 'instantiation' of it. Obviously, this natural pre-adaptation and -integration is very different from the type of pre-adaptation that Regier's system has. This point will be elaborated and discussed in further detail later since it also applies to enactive approaches.

It should be noted that the point of the discussion so far is neither that cognitivism is wrong nor that cognitivist grounding along the above lines is impossible. As Harnad (1993) pointed out, symbol grounding is an empirical issue. A cognitivist grounding theory can, however, not be considered complete as long as it only explains the causal connection of sensory percepts to individual atomic representations, but neither the transducing input system itself, nor its interdependence with its environment and the computational central systems.

Enactive Grounding

In contrast to cognitivism, the enactive framework is characterized by its focus on agent-environment mutuality and *embodied action*, which Varela et al. (1991) explain as follows:

> By using the term *embodied* we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context. By using the term *action* we mean to emphasize ... that sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. (Varela et al. 1991)

Hence, unlike traditional AI which is committed to "computer programs" (cf. Searle quote above), the preferred objects of study in enactive AI research are typically robotic agents, situated in some environment and interacting with it via sensors and motors, instead of dealing with abstract models of it. Furthermore, enactive research is based on the idea of intelligent behaviour being the outcome of the dynamical interaction of agent and environment, rather than the former's capacity to represent the latter (e.g., Varela et al. 1991; Brooks 1991a; Beer 1995). Thus, the enactive/robotic approach to AI does seem to follow Searle's 'recommendation' to focus on machines, i.e. physical systems interacting with their environments, and therefore, at a first glance, might seem immune to the grounding problem.

Physical Grounding: The key ideas of enactive AI are reflected by the commitment to "the two cornerstones of the new approach to Artificial Intelli-

framework.[1] Harnad proposed a hybrid symbolic/connectionist system in which symbolic representations (used in the central systems, in Fodorian terms) are grounded in non-symbolic representations of two types: *iconic representations*, which basically are analog transforms of sensory percepts, and *categorical representations*, which exploit sensorimotor invariants to transduce sensory percepts to *elementary symbols* (e.g. 'horse' or 'striped') from which again *complex symbolic representations* could be constructed (e.g. 'zebra' = 'horse' + 'striped'). As a natural 'candidate component' for this bottom-up transduction (from real world objects via non-symbolic representations onto atomic symbolic representations) Harnad mentioned connectionist networks (1990, 1993).

A number of approaches to grounding have followed similar lines as those proposed by Harnad. Some of them, however, deny the need of symbolic representations (e.g., see Lakoff's (1993) interpretation/evaluation of Regier's (1992) work), and accordingly transduce sensory percepts onto non-symbolic (typically connectionist) representations. For a detailed account of the differences between symbolic and connectionist computational engines and grounding approaches see (Sharkey & Jackson 1996). The symbolic/connectionist distinction will not be further elaborated in this paper, since the more relevant distinction here is that between cognitivism and enaction (connectionist approaches can be found on both sides), and the associated types of representation (explicit world models and manipulable representations vs. behaviour-generating patterns).

Although Harnad's grounding theory is based on a *robotic functionalism* (1989, 1995) rather than pure cognitivism, and he has repeatedly pointed out (1993, 1995) that categorical invariants have to be grounded in robotic capacity, i.e. in *sensorimotor* interaction with the environment, most cognitivist approaches follow the tradition of neglecting action and attempt to ground internal representations in *sensory* invariants alone. Hence, most of these approaches aim at grounding object categories (and thereby the crucial atomic representations) in perception.

[1] In fact Harnad's symbol grounding proposal has been referred to as "a face-saving enterprise" (Sharkey & Jackson 1996) for symbolic theories of mind.

Regier's Perceptually Grounded Semantics: A typical example is the work of Regier (1992) (see also Lakoff's (1993) and Harnad's (1993) discussion of Regier's work), who trained structured connectionist networks to label sequences of two-dimensional scenes, each containing a landmark and an object, with appropriate spatial terms expressing the spatial relation of the two (e.g. 'on', 'into', etc.). Or, in Regier's words: "the model learns perceptually grounded semantics".

Another example is the work by Cottrell et al. (1990) who trained connectionist networks (a) to label visual images (associate faces with names), and (b) to associate simple sequences of visual images with simple sentences.

This transduction of percepts onto manipulable internal representations, could be argued to solve the problem of representation grounding (at least partly) since it does offer a pathway from real world objects to internal representations, thereby grounding the latter.

Let us have a closer look at Regier's system though (very similar observations can be made in the case of Cottrell et al. (1990)). Do we have a truly grounded system here, i.e. is what the system does, and how, intrinsic and meaningful to the system itself? Well, of course it is not. Anything that goes on in the system, except for the produced labels, is still completely ungrounded: The system has no concept of what it is doing or what to use the produced labels for, i.e. it is not embedded in any context that would allow/require it to make any meaningful use of these labels. That means, for Regier's system to be considered to capture/possess intrinsic meaning, there are at least two things missing, which will be discussed in the following.

Firstly, the created labels (i.e. the results of the transduction) could possibly be considered grounded (see however Harnad's (1993) argument that a feature detector alone cannot provide semantics). The act of labelling (transduction) itself, however, since it does not have any functional value for the labelling system, sure cannot be considered intrinsic or meaningful to itself. That means, a semantic interpretation of the system's behaviour is of course possible ("this system labels spatial scenes"), it is, however, definitely not intrinsic to the system itself, it is just parasitic on the interpretation in our (i.e. the observers') heads.

Hence, for a system's *behaviour*, whatever it is the system does, to be *intrinsically meaningful* to

al. (1991) and taken in similar form by, e.g., Clark (1997). It should however be noted that the enactive paradigm (although, so far, relatively few researchers actually use the term 'enaction') is to a large extent compatible with constructivist views such as Piaget's genetic epistemology (cf. Rutkowska 1996), the dynamical hypothesis in cognitive science (e.g. van Gelder 1995, 1998; Port & van Gelder 1995), as well as much of the recent work on situated/embodied/behaviour-based AI and cognitive science, artificial life, autonomous agents research, cognitive robotics, etc. (cf. Varela et al. 1991; Brooks 1991b; Clark 1997; Pfeifer & Scheier 1998).

Cognitivism vs. Enaction

Cognitivism, as exemplified by the aforementioned PSSH, can be said to be "dominated by a 'between the ears', centralized and disembodied focus on the mind" (Rutkowska 1996). In particular, cognitivism is based on the traditional notion of *representationalism* (Fodor 1981; Fodor & Pylyshyn 1988), characterized by the assumption of a stable relation between manipulable agent-internal representations ('knowledge') and agent-external entities in a pregiven external world (cf. Peschl 1996). Hence, the cognitivist notion of cognition is that of computational, i.e. formally defined and implementation-independent, processes manipulating the above representational knowledge internally.

The *enaction* paradigm (Varela et al. 1991) on the other hand, emphasizes the relevance of action, embodiment and agent-environment mutuality. Thus, in the enactive framework, cognition is not considered an abstract agent-internal process, but the outcome of the dynamical interaction between agent and environment and their mutual specification during the course of evolution and the individual's lifetime. Hence, the enactive approach

> ... provides a view of cognitive capacities as inextricably linked to histories that are *lived*, much like paths that only exist as they are laid down in walking. Consequently, cognition is no longer seen as problem solving on the basis of representations; instead, cognition in its most encompassing sense consists in the enactment or bringing forth of a world by a viable history of structural coupling. (Varela et al. 1991)

This de-emphasis of representation in the traditional sense, in particular Brooks' (1991a) paper "Intelligence without Representation", has often been interpreted as denying any need for representation. There has, however, been much discussion recently of the notion of representations as "behaviour-generating patterns" (Peschl 1996) without a stable relation to environmental entities (cf. also Globus 1992; Clark & Wheeler 1998), as well as the notion of 'indexical-functional' or 'deictic' representations (e.g. Agre & Chapman 1987, Brooks 1991b), i.e. representations of entities in terms of the their functional or spatial relation to the agent, as well interactivist (Bickhard & Terveen 1995) or experiential accounts (Sharkey 1997) of representation as something constructed by an agent in interaction with an environment. All of these fit well into the enactive framework of cognition as agent-environment interaction which is thus

> ... quite compatible with viewing representation in terms of mechanisms that establish *selective correspondence* with the environment, rather than as internal models that substitute for things in the world in the overlayed traditional sense of re-presentation. (Rutkowska 1996)

Cognitivist Grounding

Typical for the cognitivist paradigm is a perception-cognition distinction (cf., e.g., Rutkowska 1996), such as Fodor's (1980, 1983) distinction into *input systems* (e.g., low-level visual and auditory perception) and *central systems* (e.g., thought and problem solving). Input systems are typically considered responsible for transducing percepts onto internal representations, whereas the central systems manipulate/reason with the representational model/ knowledge in a purely computational fashion.

Grounding Atomic Representations: In general, cognitivist grounding approaches typically focus on input systems grounding atomic representations in sensory/sensorimotor invariants. That means, here the required causal connection between agent and environment is made by hooking atomic internal representations to external entities or object categories. Such grounded atomic representations are then considered to be the building blocks from which complex representational expressions ('inheriting' the grounding of their constituents) can be constructed and a coherent representational world model can be built.

Harnad's Proposal: Harnad (1990) himself suggested a possible solution to the symbol grounding problem which mostly fits into the cognitivist

foundation/cornerstone of classical AI and cognitivism.

In particular Searle considered work by Schank and Abelson (1977), who claimed their computer programs, using so-called 'scripts', a symbolic knowledge representation technique, to be models of human natural language story understanding. To validate these claims Searle suggested a thought experiment: Imagine a person sitting in a room, who is passed (e.g., under the door) sequences of, to him/her meaningless, symbols. The person processes these symbols according to formal rules which are given in his/her native language (e.g., written on the room's walls), and returns a sequence of resulting symbols. As Searle pointed out, the symbols could, unknown to the person in the room, in fact be a story, questions and answers in Chinese written language. Hence, Chinese-speaking observers outside the room could very well conclude that who- or whatever is processing the symbols inside the room in fact does understand Chinese (since the symbols do have meaning to the observers, and the answers returned from the room might be fully correct), whereas in reality the person in the room does of course not.

Searle therefore concluded that the computer programs of traditional AI, operating in a purely formally defined manner, similar to the person in the room, could neither be said to 'understand' what they are doing or processing, nor to be models of human story understanding. According to Searle, this is mostly due to their *lack of intentionality*, i.e. their inability to relate their arbitrary internal representations (symbols) to external objects or states of affairs. Nevertheless, Searle did not suggest to give up on the idea of intelligent machines, but in fact he concluded

> ... that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains. And that is the main reason strong AI has had little to tell us about thinking, since it has nothing to tell us about machines. By its own definition it is about programs, and programs are not machines. (Searle 1980)

Harnad (1990) basically extended and refined Searle's analysis of the problem, but also proposed a possible solution how to ground symbolic representations in behavioural interactions with the environment (cf. following section). In his formulation of the *symbol grounding problem* Harnad compared purely symbolic models of mind to the attempt to learn Chinese as a first language from a Chinese-Chinese dictionary. Accordingly, he also concluded that "cognition cannot be just symbol manipulation" since the symbols in such a model, as the symbols processed in Searle's Chinese Room, could very well be

> ... systematically *interpretable* as having meaning ... [b]ut the interpretation will not be *intrinsic* to the symbol system itself: It will be parasitic on the fact that the symbols have meaning for *us* [the observers], in exactly the same way that the meaning of the symbols in a book are not intrinsic, but derive from the meaning in our heads. (Harnad 1990)

Several authors have pointed out that the grounding problem is not limited to symbolic representations, and therefore referred to it more generally as the problem of *representation grounding* (Chalmers 1992) or *concept grounding* (Dorffner & Prem 1993), or the *internalist trap* (Sharkey & Jackson 1994).

A number of approaches to grounding have been proposed, all of which basically agree in two points. Firstly, escaping the internalist trap has to be considered "crucial to the development of truly intelligent behaviour" in artefacts (Law & Miikkulainen 1994). This is very much in line with much recent research on situated and embodied AI/cognitive science (e.g., Agre & Chapman 1987; Maturana & Varela 1987; Varela et al. 1991; Brooks 1991b; Wilson 1991; Clark 1997) which considers agent-environment interaction, rather than disembodied problem solving, to be the core of cognition and intelligent behaviour. Secondly, to achieve grounding agents have to be "hooked" (Sharkey & Jackson 1996) to the external world in some way. That means there have to be causal connections, which allow the artificial agent's internal mechanisms to interact with their environment directly and without being mediated by an external observer.

Approaches to Grounding

The question of what exactly has to be hooked to what and how, however, divides the different approaches, as will be discussed in this section. For the purpose of this paper different approaches to grounding can be categorized into two groups according to whether they follow the cognitivist or the enaction paradigm in cognitive science. This rough distinction basically follows that made by Varela et

Rethinking Grounding[0]

Tom Ziemke
Department of Computer Science, University of Skövde, Box 408, 54128 Skövde, Sweden
tom@ida.his.se

[0] To appear in Riegler, Peschl, von Stein (Eds.) *Understanding Representation in the Cognitive Sciences*. New York: Plenum Press, 1999.

Version: 10 December 1998. (NB. Page numbers will change.)

NB. This is a substantially revised and extended version of a paper with the same title that appeared in Riegler & Peschl (Ed.s.) *Does Representation Need Reality?* , pp. 87-94. Austrian Society for Cognitive Science Technical Report 97-01. Vienna, Austria, May 1997.

Abstract

The grounding problem is, generally speaking, the problem of how to embed an artificial agent into its environment such that its behaviour, as well as the mechanisms, representations, etc. underlying it, can be intrinsic and meaningful to the agent itself, rather than dependent on an external designer or observer. This paper briefly reviews Searle's and Harnad's analyses of the grounding problem, and then evaluates cognitivist and enactive approaches to overcoming it. It is argued that, although these two categories of approaches differ in their nature and the problems they have to face, both, so far, fall short of solving the grounding problem for similar reasons. Further it is concluded that the reason the problem is still somewhat underestimated lies in the fact that modern situated and embodied AI, despite its emphasis of agent-environment interaction, still fails to fully acknowledge the historically rooted integrated nature of living organisms and their environmental embedding.

Introduction

The *grounding problem* is, generally speaking, the problem of how to causally connect an artificial agent with its environment such that the agent's behaviour, as well as the mechanisms, representations, etc. underlying it, can be intrinsic and meaningful to itself, rather than dependent on an external designer or observer. It is, for example, rather obvious that your thoughts are in fact intrinsic to yourself, whereas the operation and internal representations of a pocket calculator are extrinsic, ungrounded and meaningless to the calculator itself, i.e. their meaning is parasitic on their interpretation through an external observer/user. Nevertheless, the fact that the lack of grounding poses a serious problem for synthesis and modelling of intelligent behaviour in artefacts has been somewhat underestimated, not to say ignored, in the fields of artificial intelligence (AI) and cognitive science for a long time. Recent interest in the issue has been mainly triggered by the arguments of Searle (1980) and Harnad (1990).

The following section will briefly recapitulate Searle's and Harnad's formulations of the grounding problem. Different approaches to overcome the problem are then reviewed, in particular cognitivist approaches to grounding meaning in perception and enactive approaches counting on the physical grounding of embodied and situated agents. It will be argued that none of these approaches offers a satisfactory solution to the grounding problem since all of them address only part of the problem. The notion of radical bottom-up grounding of complete agents, through co-evolution/-development of (robotic) bodies, nervous systems and environments, will then be discussed as a possible route towards the development of truly grounded or *rooted* artefacts, i.e. systems whose behaviour and underlying mechanisms are in fact intrinsic to themselves, and which form a systematic, meaningful whole with their environment.

The Grounding Problem

In 1980 Searle put forward his *Chinese Room Argument* (CRA) in order to contradict the notion (which he referred to as 'strong AI') of intelligent behaviour being the outcome of purely computational, i.e. formally defined and implementation-independent, processes in physical symbol systems, as put forward in the Physical Symbol Systems Hypothesis (PSSH) (Newell & Simon 1976; Newell 1980), the