coneural
Center for Cognitive and Neural Studies

Autonomous artificial intelligent agents

Răzvan V. Florian

Center for Cognitive and Neural Studies (Coneural)
Str. Saturn 24, 3400 Cluj-Napoca, Romania
www.coneural.org
florian@coneural.org

Technical Report Coneural-03-01
February 4, 2003

Abstract

This paper reviews the current state of the art in the research concerning the development of autonomous artificial intelligent agents. First, the meaning of specific terms, like agency, automaticity, autonomy, embodiment, situatedness, and intelligence, are discussed in the context of this domain. The motivations for conducting research in this area are then exposed. We focus, in particular, on the importance of autonomous embodied agents as support for genuine artificial intelligence. Several principles that should guide autonomous agent research are reviewed. Of particular importance are the embodiment and situatedness of the agent, the principle of sensorimotor coordination, and the need for epigenetic development and learning capabilities. They ensure the adaptability, flexibility and robustness of the agent. Several design and evaluation considerations are then discussed. Four approaches to the design of autonomous agents—the subsumption architecture, evolutionary methods, biologically-inspired methods and collective approaches—are presented and illustrated with examples. Finally, a brief discussion mentions the possible role of autonomous agents as a framework for the study of computational applications of the far-from-equilibrium systems theory.

Contents

1 Introduction

Autonomous intelligent agent research is a domain situated at the forefront of artificial intelligence. As shown below, it was argued that genuine intelligence can emerge only in embodied, situated cognitive agents. It is a highly interdisciplinary research area, connecting results from theoretical cognitive science, neural networks, evolutionary computation, neuroscience, and engineering. Besides its scientific importance, there are also important applications of this domain in the development of robots used in industry, defense and entertainment.

We will first attempt to delimit the scope covered by the term "autonomous artificial agent". The scientific importance of the study of embodied agents will then be stressed. The paper will continue with the presentation of the principles used in the design of artificial autonomous agents. Design and evaluation considerations will be also discussed. Several design methods will be then illustrated. Finally, we will briefly discuss the possible role of autonomous agents as a framework for the study of computational applications of the theory of far-from-equilibrium systems.

2 What is an autonomous intelligent agent?

Agency, autonomy, and intelligence are notions that are all fuzzy and hard to define. Also, agency is tightly connected to qualities like autonomy, situatedness, and embodiment. Most authors refrain to give precise definitions, as such definitions are inevitably either too extended or too narrow. For example, Russell and Norvig (1995) consider: "The notion of an agent is meant to be a tool for analyzing systems, not an absolute characterization that divides the world into agents and non-agents." Moreover, the different definitions available in the literature are often not consistent with the others.

Without attempting to explain precisely these terms, we will outline here their meaning, in order to delineate the scope of this paper.

2.1 Agency, automaticity, autonomy

We generally consider humans and most other animals as being agents. Scientists and engineers have also built robots, systems and software programs that can be considered to be artificial agents. But what really distinguishes an agent from other artificial systems?

Luc Steels (1995), a preeminent artificial intelligence researcher, considers that the essence of *agency* is that "an agent can control to some extent its own destiny". This requires *automaticity*—the agent to have mechanisms that allow the agent to sense the environment and act upon it and do not require the intervention of other agents to be executed. A thermostat or a virus can be thus considered to be an agent.

Autonomy is a characteristic that enhances the viability of an agent in a dynamic environment. For autonomous agents, "the basis of self-steering originates (at least partly) from the agent's own capacity to form and adapt its principles of behavior. Moreover, the process of building up or adapting competence is something that takes place while the agent is operating in the environment" (Steels, 1995). Autonomy requires automaticity, but goes beyond it, implying some adaptability. However, autonomy is a matter of degree, not a clear cut property (Smithers, 1995; Steels, 1995). Most animals and some robots can be considered autonomous agents.

Other authors consider that agents are implicitly autonomous. In a study seeking to draw the distinction between software agents and other software systems, Franklin and Graesser (1996) have made a short survey of the meaning of "agent" in the computer science and artificial intelligence literature. In the papers surveyed there, agency is considered to be inseparable from autonomy.

As a conclusion their survey, Franklin and Graesser attempt a definition: "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future."

Ordinary computer applications, such as an accounting program, could be considered to sense the world via their input and act on it via their output, but they are considered not to be agents because their output would not normally effect on what it senses later. All software agents are computer programs, but not all programs are agents (Franklin & Graesser, 1996).

Agents are different from the objects in object-oriented computer programs by their autonomy and flexibility, and by having their own control structure. They are also different from the expert systems of classical artificial intelligence by interacting directly with an environment, and not just processing human-provided symbols, and also by their autonomous learning (Iantovics & Dumitrescu, in press).

Pattie Maes from MIT Media Lab, one of the pioneers of agent research, also defines artificial autonomous agents (Maes, 1995), as "computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed."

2.2 Situatedness

From the definitions above, *situatedness*—the quality of a system of being situated in an environment and interacting with it—seems to be regarded as an implicit property of most agents.

2.3 Embodiment

Embodiment is an important quality of many autonomous agents. It refers to their property of having a body that interacts with the surrounding environment. This property is important for their cognitive capabilities, as we will see below. While this generally refers to a real physical body, like those of animals and robots, several studies (Quick, Dautenhahn, Nehaniv, & Roberts, 1999; Riegler, 2002; Oka et al., 2001) have argued that the importance of embodiment is not necessarily given by materiality, but by its special dynamic relation with the environment. A body can both be influenced by the environment and act on it. Some of its actions can change the environment, thus changing the influence of the environment over it, in a closed loop structural coupling. This can also happen in environments other than the material world, such as computational ones. The environment can be a simulated physical environment, or a genuinely computational one, such as the internet or an operating system. Embodiment is thus defined extendedly by Quick et al. (1999): "A system X is embodied in an environment E if perturbatory channels exist between the two. That is, X is embodied in E if for every time t at which both X and E exist, some subset of E's possible states have the capacity to perturb X's state, and some subset of X's possible states have the capacity to perturb E's state." This is closely related to the biologically inspired idea of structural coupling from the work of Maturana and Varela (1987).

Ziemke (2001a, 2001b) also discusses other forms of embodiment: "organismoid" embodiment, i.e. organism-like bodily form (e.g., humanoid robots), and the organismic embodiment of autopoietic, living systems. He also notes that embodiment may be considered a historical quality, in the sense that systems may not only be structurally coupled to their environment in the present, but that their embodiment is in fact a result or reflection of a history of agent-environment interaction. In our interpretation of the term, embodiment must be historical, but not necessarily organismoid nor organismic.

Embodiment is tightly connected to situatedness—a body is not sufficient for embodiment, if it is not situated in an environment. Moreover, the body must be adapted to the environment, in order to have a mutual interaction. In this interpretation, a robot standing idle on a shelf, a robot having only visual sensors but inhabiting an environment without light, or a robot which does not perceive its environment, acting according to a predefined plan or remotely controlled, are not considered to be embodied nor situated.

2.4 Intelligence

Intelligence is another hard to define notion, and even a controversial one. Various authors consider it to be an ability to learn from experience, to adapt to new situations and changes in the environment, or to carry on abstract thinking (Pfeifer & Scheier, 1999). The MIT Encyclopedia of Cognitive Science states: "An intelligent agent is a device that interacts with its environment in flexible, goal-directed ways, recognizing important states of the environment and acting to achieve desired results" (Rosenschein, 1999).

However, in practice intelligence is a relative attribute, and evaluated in connection with human capabilities. For example, we would not normally consider a rat being intelligent (implying a comparison to a human), but we would recognize it to be more intelligent than a cockroach.

In agreement to these considerations, in this paper we would consider an agent as being intelligent if it is capable of performing non-trivial, purposeful behavior that adapts to changes in the environment. However, the evaluation of the behavior is arbitrarily done by a human and thus intelligence is a subjectively assigned property.

3 Reasons for studying artificial autonomous agents

3.1 Applications

Many artificial agents are developed for performing physical tasks that directly serve human purposes. Scientists and engineers are trying to build robots that can relieve people of dangerous, physically demanding, or monotonous jobs. Many robots automate work in the manufacturing industries; however, they are usually not autonomous nor intelligent. Other robots, with various degrees of autonomy, are used for exploring remote or inaccessible locations. For example, they might investigate distant planets (like the Mars Sojourner[1]) or the ocean floor, they might inspect oil pipelines (like iRobot's MicroRig system[2]) or sewer pipes (like MAKRO; Kolesnik & Streich, 2002). Their autonomy may eliminate the need for expensive remote control equipment (like kilometers of cable and machines for manipulating the cable, in the case of pipe or sewer inspection), or for human surveillance operators. Autonomy may also protect them in the case of unexpected events, when the remote controlling operator is not capable to respond fast enough to these events due to delays in communication, like in planetary exploration. Research is being carried for creating robots that can rescue people from crushed buildings or for demining operations.

[1] http://mars.jpl.nasa.gov/MPF/rover/sojourner.html
[2] http://www.irobot.com/industrial/microrig.asp

Consumer robotics is estimated as a huge market, especially in the context of the increasing number of aged people in the developed countries. iRobot's Roomba[3], launched in 2002, is the first consumer automatic robotic vacuum cleaner.

Artificial intelligent agents are also used for entertainment, as virtual companions or in movies and graphics. For example, the computer game Creatures[4] features artificial characters that grow, learn from the user, and develop their own personality. The Sony Aibo robotic dog[5] behaves like an artificial pet, entertaining their owners and even emotionally attaching to them, through their interactive behavior. Artificially evolved neural network controllers for computer simulated fish were used for generating realistic computer graphics (Terzopoulos, 1999).

There are thus important possible applications for autonomous intelligent agents. However, the degree of autonomy and intelligence of current artificial agents is quite low, in comparison with biological ones, like mammals. Research is being carried for improving the autonomy and intelligence of artificial agents. This paper will present next some principles, many of which are biologically inspired, that should be followed for developing more competent artificial intelligent agents (Section 4).

3.2 Autonomous agents as support for genuine artificial intelligence

Autonomous agents research is not only interesting for its immediate applications for physical tasks, but also for the more general purpose of developing genuine artificial intelligence. As we will show next, it is currently considered that genuine intelligence can emerge only in situated, embodied agents, which can interact directly with an environment.

3.2.1 Classical artificial intelligence

At the beginning of these disciplines, starting with the 50's, most of the researchers in artificial intelligence (AI) and even in cognitive science, in general, considered reasoning a disembodied process. These first years of cognitive studies were particularly marked by the influence of the computer, that was a relatively new technology at that time. Intelligent behavior was often viewed as computation. It was thought that human intelligence is achieved by symbolizing external and internal situations and events and by manipulating these symbols according to syntactic rules (Fodor, 1975; Pylyshyn, 1980; Simon & Kaplan, 1989). The supporters of this so-called cognitivist or functionalist approach sustained that once the good algorithms

[3] http://www.roombavac.com
[4] http://www.creatures3.com
[5] http://www.aibo.com

and ways of representing knowledge in symbols would be found, intelligence can be implemented in any kind of computing machines, like computer software, regardless of the hardware implementation. In this framework, the body of the cognitive agent is not regarded to have a particular relevance: it may provide symbolic information for input, or act out the result of the computation, like a peripheral device, or it may be lacking at all. The only important process is considered to be the symbol manipulation in the central processing unit.

Until the 80's, most of the models in cognitive science and cognitive psychology were inspired by the functioning of the computer and phrased in computer science and information processing terminology; some of these models continue to be backed today by their supporters. Representational structures such as feature lists, schemata and frames (knowledge structures that contain fixed structural information, with slots that accept a range of values), semantic networks (lists and trees representing connections between words) and production systems (a set of condition-action pairs used as rules in the execution of actions) were used to explain and simulate on computers cognitive processes (Anderson, 1993; Newell, 1990). It was proposed that problem solving is accomplished by humans through representing achievable situations in a branching tree and then searching in this problem space (Newell & Simon, 1972). It was also proposed that objects are recognized by by analysis of discrete features or by decomposing them in primitive geometrical components (Biederman, 1987).

In robotics, the efforts were directed towards building internal models of the world, on which the program could operate to produce a plan of action for the robot. Perception, planning and action were performed serially. Perception updated the state of the internal model, which was predefined by the designer of the robot. Because of this, perception recovers predetermined properties of the environment, rather than exploring it. The environments in which robots evolved were often fixed, otherwise the internal model would have failed to represent reality. Planning was achieved through symbolic manipulation in the internal world model. A classical example of this sense-model-plan-act (SMPA) approach (Brooks, 1995, p. 28) is the robot Shakey built in the 60's at the Stanford Research Institute[6].

3.2.2 Limits of classical AI

The methods of this so-called Good Old Fashioned Artificial Intelligence (GOFAI) had some impressive successes in certain domains; however, these successes are limited. Based on those methods, programs were built that solved problems and proved theorems from logic and geometry. However, they depend on humans for converting the problem in a representation suitable for them and are confined to domains where knowledge can be easily

[6] http://www.sri.com/technology/shakey.html

formalized. Expert systems are widely used in the industry for the planning of the processes, but once the situation gets out of their ontology, they have no capability of dealing with it. One of the best known expert systems is MYCIN (Shortliffe, 1976), a program for advising physicians on treating bacterial infections of the blood and meningitis. An example of MYCIN's limitations is to tell MYCIN that that Cholerae Vibrio was detected in the patient's intestines. The system will recommend two weeks of tetracycline and nothing else. This would probably kill the bacteria, but most likely the patient will be dead of cholera long before the two weeks. However, the physician will presumably know that the diarrhea has to be treated as well (McCarthy, 1984; see also Brooks, 1991).

The defeat of the world chess champion, Garry Kasparov, by the Deep Blue computer in 1997 was widely publicized[7]. Another expert system, a recent computer program built on the FORR architecture (Epstein, 1994) is capable of learning and successfully playing several types of games. However, a program that would beat a professional go player is still yet to be built, because the search space is much bigger in go than in other games[8]. This is a good example where traditional methods fail.

Research in Natural Language Processing have led to programs that are able to search and summarize text, to translate automatically, and to chat with a human partner. The state of the art programs in this field can be easily tested on the web[9]: neither the word-by-word translation, nor the grammatical analysis of the phrase structure are enough to understand natural language. These problems point to the fact that understanding of the semantics and information about the context are crucial. The commercial Cyc project[10], still under development, struggled for more than ten years to build a huge semantic net that would cover the commonsense knowledge of an ordinary human. In spite of the huge quantity of information fed into computers, the results are well below expectations.

In general, most intuitive human knowledge still resists formalization, including that involved in comprehending simple stories or simple physical situations. Our surrounding environment has a much too complex structure, that cannot be captured by a single ontology. This follows not only from theoretical considerations (Popper, 1959), but was also shown by modern physics (Feynman, 1965 / 1992, chap. 7). Classical AI systems are usually brittle, in the sense that they are unable to adapt to situations unforeseen by their programmer, to generalize, and lack tolerance to noise and fault tolerance. Their preprogrammed nature prevents them to display creative behavior. Many day-to-day human problems seem to have unmanageable

[7] http://www.chess.ibm.com/

[8] http://www.intelligentgo.org/en/computer-go/overview.html

[9] Babelfish, automatic translator: http://babelfish.altavista.com; chat bots: http://www.botspot.com/search/s-chat.htm

[10] http://www.cyc.com

computational complexities for systems designed in the framework of classic artificial intelligence. There is little direct evidence that symbol systems underlie human cognition (Barsalou, 1999), although it was proposed at times that the human brain functions under similar principles.

3.2.3 Fundamental problems of classical AI

Besides their lack of biological plausibility and the practical problems in implementing them as intelligent systems, disembodied symbol systems exhibit more fundamental problems.

The symbol grounding problem (Harnad, 1990) refers to the fact that, in classical symbolic systems, there is nothing to give meaning to the manipulated symbols, for the systems that perform this manipulation. These symbols have meaning (representational content) for external observers—the human designer, programmer or user of these systems, but not for the systems themselves. Bickhard (1993) argues on theoretical grounds that genuine representational content can emerge only in an embodied, goal directed agent, that is able to perceive its environment and interact with it. Representation of different environmental situations emerges if the agent is able to distinguish different potentialities for action in these situations, related to its goal. This theoretical framework is substantiated by psychological and other experimental evidence. This shows that human experience occurs when the organism masters the laws of sensorimotor contingency (O'Regan & Noe, in press), i.e. anticipates the changes in perception that may be produced by potential actions.

The frame of reference problem refers to the confusion between the terms of the description of intelligent behavior by an observer, and the real mechanism that generates this behavior; and between the perspective of an observer and the perspective of the intelligent agent itself (Pfeifer & Scheier, 1999, pp. 82, 111–117). For example, if a human observes an agent performing a certain task, this doesn't automatically imply that there is an internal representation of the task within the agent. Knowledge level descriptions constitute an observer's model, not structures or mechanisms inside the agent (Clancey, 1995, p. 228). Moreover, the segmentation of behavior by a human observer is arbitrary; and the behavior of an agent is always the result of a system-environment interaction. As a result, the observed characteristics of a particular behavior do not always indicate accurately the complexity or the nature of the underlying mechanisms.

An illustration of this fallacy may be given by some simple vehicles with two sensors and two wheels powered by independent motors, and very simple wiring between the sensors and the motors (Braitenberg vehicles; Braitenberg, 1984). If the sensors have nonlinear characteristics, these vehicles can exhibit very complex behaviors. Human observers may attribute them will or personality; however, they act according to extremely simple rules.

In neural systems, the activation of a neuron correlated with an observable feature of the environment does not necessarily mean that the neuron codes for, or represent, that feature. For example, in a classic experiment, Lehky and Sejnowski (1988) trained a network with backpropagation to extract height information from shading, as presented in pictures of smooth 3D objects. In an analysis of the resulted network, they observed neurons that reacted optimally to bars and edges, like in the mammalian primary visual cortex. However, this particular network has never experienced bars nor edges during training (Churchland & Sejnowski, 1994, pp. 183–188). Also, many evolved neural networks embedded in artificial agents that successfully perform the desired tasks are difficult to be functionally analyzed (Ruppin, 2002). This questions the attempts to understand biological neural networks in terms of representations, as many neuroscience studies still try. Simple as it is, the frame of reference problem is still ignored even today by many researchers in cognitive science.

From a designer point of view, the solution for achieving a desired behavior in an artificial agent may not necessarily be related to the terms in which the problem is described (Hallam, 1995, pp. 220–221). Many interesting behaviors of biological and artificial agents appear through emergence and self-organization (see also Section 4.3).

More detailed critiques of symbol systems and classical AI are articulated by Bickhard (1993), Barsalou (1999), Pfeifer and Scheier (1999, chap. 3), Brooks (1991), Steels and Brooks (1995).

3.2.4 Embodiment as a condition for learning and adaptability

A genuinely intelligent system should be adaptive, flexible, robust: it should adjust its operation to unexpected changes that influence it, and should be creative in finding solutions for completing its tasks. We have seen that preprogrammed symbol systems cannot acquire this flexibility. Developers cannot predict and code responses for all possible situations. The speed of current computing systems is still too slow, relative to the huge search space, for evolutive methods alone to generate generic intelligent systems (Grand, 1998). Artificial intelligent systems should then develop most of their cognitive structure by learning and self-organize to arrive at emergent new behaviors; their designers should just implement sensible learning methods.

We have seen that cognitive systems cannot understand the meaning of symbols if they are not grounded through association with sensorimotor interaction. They cannot thus be initially taught through symbolic communication with humans or other agents. As for humans, an environment may offer to artificial systems a learning framework for the development of cognitive structures.

For learning, and thus interaction with the environment to be possible, the artificial cognitive system has to be able to perceive it and influence it through effectors, and thus to be *embodied and situated.*

Studies in developmental neuroscience and robotics have shown that perception, without action, is not sufficient for the development of cognitive capabilities in animals, or for interesting performance in artificial systems. For example, in a classical experiment (Held & Hein, 1958), a group of kittens were immobilized in carriage, to which a second group of kittens, who were able to move around normally, were harnessed. Both groups shared the same visual experience, but the first group was entirely passive. When the animals were released after a few weeks of this treatment, the second group of kittens behaved normally, but those who had been carried around behaved as if they were blind: they bumped into objects and fell over edges. This study supports the idea that objects are not seen by the visual extraction of features, but rather by the visual guidance of virtual action (Varela, 1995, pp. 16-17, Robbins, 2002). Also, research in active vision (Blake & Yuille, 1992) has shown that artificial vision systems where the cameras are able to move, orient, focus, etc. give much better results than passive ones in image processing and recognition problems.

If the artificial system should exhibit self-organization and emergence of interesting features, the control part has to be a distributed, far from equilibrium system, formed by a large number of interacting subunits. Artificial neural networks (ANNs) seem to be suitable for this. The fact that biological intelligence is physically implemented in networks of neurons also encourages the use of biologically-inspired ANNs for obtaining artificial intelligence.

However, interaction with the environment may offer the cognitive agent the possibility for associating meaning to symbols, by communicating and interacting with other agents that also have sensorimotor access to the same environment (e.g. Steels, Kaplan, McIntyre, & Looveren, 2000; see also Section 4.12). In contrast to the symbol systems of classical AI, these symbols have meaning for the artificial agent themselves, and are grounded in perception and action. After the possibility for symbolic communication emerges, artificial agents may also eventually be taught like this. Teaching is also possible through imitation or physical guidance in the environment (e.g. Kozima, Nagakawa, & Yano, 2002; Andry, Gaussier, & Nadel, 2002; Alissandrakis, Nehaniv, & Dautenhahn, 2001).

3.2.5 Embodied, interactivist-constructivist cognitive science

In general, there is a convergence of results from a wide range of domains within cognitive science that point to the conclusion that intelligence can arise only in embodied agents, artificial or biological, and that embodiment and situatedness also offer a more appropriate framework for the study of human and animal intelligence. There are theoretical, philosophical and biological arguments (Bickhard, 1993; Varela, Thompson, & Rosch, 1992; Varela, 1995; Chiel & Beer, 1997; Ziemke, 2001c). In AI, Rodney Brooks, director of the MIT AI Lab proposed in the early 90's that representation based methods should be discontinued, because of the practical problems of the classical approach (Brooks, 1990, 1991). Research in "nouvelle AI" should rather deal with building complete systems implemented in robots, simple at first and then incrementally more intelligent. It was argued that building embodied cognitive agents is a promising path to attain artificial intelligence (Steels & Brooks, 1995; Pfeifer & Scheier, 1999), maybe the only one given the technological capabilities available today.

For humans, it was also argued that even abstract reasoning is grounded on sensorimotor capabilities (Barsalou, 1999; Indurkhya, 1992; Lakoff & Nunez, 2000; Florian, 2002). It is believed that imagery and short term memory share many common neural mechanisms with perception or motor action (Kosslyn & Thompson, 2000; Fuster, 1995; Jeannerod, 1994, 1999). Many results point out that the neural correlates of a certain concept, activated, for example, by a word, are activations of the neural networks that were also active during the experiences of the person with the significant of that word (Damasio, 1990; Pulvermuller, 1999; Martin, Ungerleider, & Haxby, 2000). These facts seem to confirm an interactivist-constructivist view of cognition: representations depend on the interaction of the cognitive agent with the external environment and are constructed according to his individual history of interactions.

Autonomous intelligent agent research is thus not only useful for its immediate applications in physical tasks, but also for the longer term goal of obtaining genuine artificial intelligence. Once appropriate cognitive structures emerge in embodied agents after learning, and symbolic communication can be established with them, they may eventually be disconnected from their bodies. These intelligent agents would eventually be used for solving problems in more abstract domains like engineering, design, science, management of complex processes and systems and so on.

3.3 Biological modelling

Another reason for studying autonomous artificial agents is the investigation of the principles that underlie animal or human behavior. Understanding how animals work is a problem of "reverse engineering". Rather than building something with a certain functional capability, we have something that already functions and want to figure out how it works. The application of engineering methodologies seems to be an appropriate and promising approach, though not easy to implement.

Biorobots are now enabling biologists to understand these complex animal-environment relationships. They can detect and map sensory signals at the level of the animal and can measure how the presence and motion of the animal affects those signals. This data, coupled with observations of the animal itself, can lead to very sophisticated hypotheses about what is causing a behavior and what is shaping it. These hypotheses can be tested both with the biorobot and with the animal itself. The robot offers several advantages over the real animal in such studies. The behavior under test in the robot is not affected by competing, uncontrolled, behaviors. Also, much more data can be obtained from a robot, compared to an animal, on its actions, sensory input, and internal states (Webb & Consi, 2001).

For example, robots were built for investigating cricket phonotaxis (Webb, 1994), navigation of the housefly (Franceschini, Pichon, & Blanes, 1992), ant navigation based on a polarized light compass (Lambrinos et al., 1997), lobster chemo-orientation, hexapod walking, or human joint attention behavior (Webb & Consi, 2001).

4 Design principles for autonomous agents

The design of autonomous agents is an active area of research. An established theory regarding autonomous intelligent agents does not exist. The field is relatively young, having gotten out of the influence of classical AI only in the beginning of the 90's. There exist, however, several principles that may guide the design of autonomous agents. Some of the principles presented here were articulated by Pfeifer and Scheier (1999), who captured compactly many insights that were usually implicit in the previous research literature. A few others were not on the list compiled by them, but we felt that they deserve the same status as the first ones.

These principles are rather idealistic: there exist currently no artificial agents that implement all of these principles. However, these principles may guide researchers in the quest of obtaining genuine artificial intelligence.

4.1 The three–constituents principle

Designing autonomous agents always involves three constituents: (1) definition of the ecological niche, (2) definition of desired behaviors and tasks, and (3) the design of the agent (Pfeifer & Scheier, 1999, pp. 302–306).

The range of environments that agents may inhabit can exhibit a lot of variety. No single agent can adapt, both physically and cognitively, to cope economically with all the possible variations. Biological agents, animals or plants, are also limited in their adaptability to a specific environmental niche. A desired ecological niche must thus be established, prior to the design of the agents.

Given the specific niche, the desired behaviors or tasks to be solved can be specified, and then the agent may be designed according to these needs. In some cases, the physical design of the agent is given (for example, if the robot is bought off-the-shelf), and only the control system can be designed, given the desired behaviors. In other cases, there might be a given agent architecture and the research will consist in the exploration of the emerging behaviors in a particular ecological niche.

The three constituents are interdependent: the design critically depends on the desired behaviors and the niche, the possible behaviors are dependent on the environment and the agent, and the ecological niche where the agent is viable depends on its structure and on what it does.

4.2 Autonomy, embodiment, situatedness

Ideally, autonomous agents should be able to function with little human intervention, supervision or instruction. They should be self-sufficient, i.e. able to sustain themselves over extended periods of time (Pfeifer & Scheier, 1999, p. 306). However, as we previously discussed (Section 2.1), autonomy is a graded property. In some cases, a high degree of autonomy is not necessary, if the agent is useful even if it depends on some external support. For example, even a human manager may need the permanent services of several assistants for successfully doing his job.

As previously shown (Section 3.2.4), the agent should be embodied and situated, in order to be able to adapt to the the structure of its environment and ground his cognitive structures. The body may be as important for adaptability as the control system (Chiel & Beer, 1997). The embodiment may be physical or computational (Section 2.3).

4.3 Emergence, self-organization

Emergence is a potential solution to the frame of reference problem (Section 3.2.3), and, more generally, to the problem of generating artificial intelligence. The human designer of an artificial agent would like it to be able to perform certain tasks. The designer has a certain conceptualization or description of the desired task. However, he may not always design efficiently the agent according to his view of the task. This was the approach of classical AI: formalize the problem and then implement a symbolic solution for it. As we have seen, this leads to brittle, unadapted systems. The preprogrammed agent will not be able to generate new behaviors that were not initially implemented, or vary the implemented ones. Moreover, any conceptualizations of a given process imply a simplification of the reality, so the designer may not be aware of all relevant issues, except in very simple environments. His segmentation of behavior is arbitrary. His view of the task depends on a human perspective, dependent on human goals and sensorimotor capabilities. The agent may have a different embodiment, and thus a different perspective of the task. The behavior is usually not entirely dependent of the agent's actions, but is the result of the interaction between the agent and the environment. Further, the symbolic description of the behavior needed for writing the computer program that controls the agent may result in large distortions of the intended structure, given by the constraints of the functioning of the computer and the programming conventions.

The solution to these problems is to design the system for emergence: behavior that is not preprogrammed should result from agent-environment interaction and from the self-organization of the agent's control system. Several principles that give some hints about how this can be accomplished are presented next.

4.4 Epigenesis, online learning

Epigenesis is a special case of emergence: it is a process through which increasingly more complex cognitive structures emerge in a system as a result of interactions with the physical and social environment. The term was introduced in psychology by Jean Piaget, to refer to such development, determined primarily by the interaction between the organism and the environment, rather than by genes. Psychology still provides empirical findings and theoretical generalizations that may guide the implementation of artificial systems capable of epigenesis (Zlatev & Balkenius, 2001).

There are two important characteristics of epigenesis that must be highlighted. First, the role of the environmental factors is constructive rather than being only selective. Many other approaches to the developmental interaction between an agent and environment stress the role of specific input either in permitting a developmental process to unfold, or in parametrically selecting a particular variant of development. In neither of these cases does the environmental information add any higher level of organization to the existing cognitive structures of the agent. The pathway along which the behavior develops, and its terminal structure, are assumed in these approaches to be predetermined. By contrast, in epigenesis the developmental pathway and final structure of the behavior that develops are a consequence of both environmental information and existing information. For example, the development of birdsong seems to involve reproduction by imitative learning rather than selection from amongst pre-established alternatives. Fledglings not exposed to a model do develop birdsong, but it is impoverished or unelaborated relative to that of those individuals developing in a normal environment in which models are available. The second key characteristic of epigenesis is that an initially specified developmental envelope or window specifies an initial behavioral (or perceptual) repertoire that is subsequently elaborated through experience of a relevant environment (Sinha, 2001).

Enabling artificial agents to epigenetically develop their cognitive structures may solve the previously mentioned problems of preprogrammed systems. In this paradigm, the designer of the artificial agent would not have to program it for specific tasks. A developmental system must be able to learn tasks that its designers do not know or even cannot predict. New tasks and skills would be learned without requiring a redesign of the control system. To design the control system of the artificial agent, the designer needs only information about the ecological niche of the agent and about its body. The designer should focus on self-organization schemes, rather than task-specific algorithms. Human teachers may affect the developing agent only as a part of the environment, preferably without interfering with its internal representation. Training may be performed by reinforcement learning, imitation or guidance (Weng et al., 2001; Weng & Zhang, 2002).

Like for humans and animals, artificial agent learning should be "online", in real time, not necessarily separated from actual performance. It should not be limited to pre-specified learning epochs, but continue during the lifetime of the agent. This would ensure that the agent will adapt in real time to unexpected changes in the environment, whenever they may arise.

It is also hypothesized that limitations of the sensory and motor systems, or of the control system, early in the developmental process of the agents, may make the learning tasks more tractable. The initially immature resources may facilitate, or even enable, the early stages of learning. Such initial limitations, followed by maturation, are common in animals. Several studies regarding learning in neural networks and robots seem to confirm that this idea is also valid for artificial systems (Lungarella & Berthouze, 2002; Clark & Thornton, 1997).

Several examples of agents built according to epigenetic principles are given by Balkenius, Zlatev, Kozima, Dautenhahn, and Breazeal (2001), Prince, Demiris, Marom, Kozima, and Balkenius (2002), Pfeifer et al. (2001, 2001).

4.5 Parallel, loosely coupled processes

Implementing the control system as a collection of parallel, heterogenous, loosely coupled processes is another principle that constitutes a support for emergence. The processes run asynchronously and are coupled to the agent's sensorimotor apparatus, requiring little or no centralized resources. An explicit process that controls all the others is unnecessary. The control is decentralized and distributed. Intelligent behavior may emerge from the joint dynamics of a number of basic processes, each of which contributes to the overall function, as the agent interacts with the environment. The architecture of the control system may develop gradually, with new processes being added on top of the others, as in biological evolution (Pfeifer & Scheier, 1999, chap. 11).

The brain itself is a massively parallel system, giving support to this principle. Artificial neural networks are also parallel systems, which are successfully used for the control of autonomous agents. Another implementation of this principle is the subsumption, behavior-based architecture (Brooks, 1986), widely used currently in robotic control, or similar ones.

The subsumption architecture will be presented in more detail below (Section 7.1). Systems constituted from many individual agents, where collective behavior emerges from local interactions, may be also seen as an implementation of this principle. Examples of such systems may be societies of ants or termites; artificial multiagent systems were also built and will be also discussed below in more detail (Section 7.4).

This principle contrasts to the centralized, sequential approach of classical AI. The classical systems are not fault tolerant and not robust with respect to noise: when a module is removed or breaks down, the whole system's functionality will be affected, because of the serial processing. Classical hierarchical architectures also prevent emergence. Parallel systems are not affected by these problems.

4.6 Sensorimotor coordination

Perception and action should always be coordinated in artificial agents, as they are in animals and humans. Their temporal separation in separate stages, as in the sense-think-act cycle of classical AI, is artificial and prevents the potential emergence of adaptive behaviors from their coordination. In biological agents, there exists a permanent dynamical recurrent interaction between perception and action, in relation with the environment and the body and the control system of the agent. Perception guides action, in interaction with the internal state of the agent. Action may change the internal state of the agent, and thus influence the internal dynamics induced by future sensations. Action also changes the perspective of the environment, as perceived by the agent, influencing thus through the environment the future sensations.

Cognition, and especially learning, needs not only perceptual, but also effector capabilities. Several empirical arguments were presented in Section 3.2.4. In another perspective, it is this sensorimotor coupling, mediated by the body and the environment, that constructs the cognitive structures (Varela, 1995, pp. 15-16). As seen above, representation and experience arise out of potentialities for action, innate (discovered through evolution) or discovered from previous experiences within the environment (Bickhard, 1993; O'Regan & Noe, in press). The cortical substrate of memory is identical to the connective cortical substrate that sustains perception and action (Fuster, 1995). Associations between past sensations and actions may offer a mechanism for anticipating the results of future actions through internal simulation (Hesslow, 2002) and thus for planning.

It was shown, both in animals and with artificial agents, that action is very important for categorization and recognition, processes traditionally conceived as purely perceptual. Mechanisms of sensorimotor coordination can be used to transform, or re-represent, information structures that are impossible to predict by means of statistical learning procedures (having hidden or marginal regularities) to learnable, predictable information structures (Pfeifer & Scheier, 1999, chap. 12, Clark & Thornton, 1997).

4.7 Goal directedness

The agent must be goal directed, must have a value system that would guide its behavior. A value system also modulates the learning process, either explicitly or implicitly. In an explicit value system, value signals that modulate learning are generated as consequences of behavior. In an implicit value system, modulation is achieved by mechanisms that select interactions with the environment, which will influence the development of the agent through learning (Pfeifer & Scheier, 1999, chap. 14). Goal directedness is also a key ingredient for the emergence of representation (Bickhard, 1993). In biological agents, the implicit goal is the survival of the species and self-maintenance. Artificial agents are usually built to perform tasks for the benefit of human users.

In order to start doing something once its starts its existence, the agent must have some innate (predefined) drives or reflexes that would induce an exploration of the environment. This exploration may lead to non-trivial sensorimotor patterns, and as a consequence to self-organization (unsupervised learning). The novel behaviors that emerge may lead to further exploration of the environment.

The self-organizational processes that lead to adaptive behavior should be reinforced. Some reinforcement may be predefined by the designer of the agent. Other reinforcement signals may be delivered to the internal control mechanisms by a user of the system, especially in the early part of the agent's interaction with the environment, if this would not prejudice the long term autonomy of the agent. Reinforcement cues may be also delivered through the environment, if the agent can relate the cues to its internal reinforcement system. More supervised learning schemes should be avoided, as they impose to the agent a human ontology, which finally prevents the agent's adaptability. The agent should rather develop its own ontology of the environment, grounded in the sensorimotor interaction. Reinforcers should just guide the interaction with the environment, not specify it.

There is a trade-off between specificity and generality of value systems. If value systems are too specific, the system is not sufficiently flexible: it is unable to generate behavioral diversity, which may be needed for attaining a goal in a complex environment. If value systems are too general, they are of little selectional value and insufficiently constrain the very large space of possible actions (Pfeifer & Scheier, 1999, chap. 14).

The goal of an agent, as interpreted by an observer of the agent's behavior, may not necessarily be the goal that the agent is following implicitly.

4.8 Cheap design

Good designs are cheap and parsimonious. The physics of the agent-environment interaction and the constraints of the ecological niche should be exploited, where possible.

An example that illustrates this principle may be a moving robot with inertia. If it is moving fast, it should turn earlier to avoid obstacles than if moving slowly, in order to minimize the risk of collisions. Intuitively one would think that this would require an assessment of the robot's speed and distance from the obstacle, and a mechanism to adjust the distance at which the agent should begin avoidance action. This turns out to be unnecessary if motion detection is employed instead, as by flies, for example. If collision detection is based by optic flow (the angular speed relative to the eye or camera), no internal mechanism for determining speed is needed (Pfeifer & Scheier, 1999, pp. 435–445).

4.9 Redundancy

An agent has to incorporate redundancy. A redundant control system is fault tolerant. An implementation as parallel, loosely coupled processes can easily assimilate redundancy. In relation to the sensory system, the principle states that an agent's sensors should have types and positions in such a way that there is potential overlap in the information that can be acquired from the different sensory channels. Correlations and associations between the inputs of different modalities may lead the agent to learn to predict sensory inputs, or to reduce uncertainty. Correlations can arise because of temporal coincidence, or may be generated through sensorimotor coordination (Pfeifer & Scheier, 1999, pp. 446–455).

4.10 Ecological balance

There has to be a balance of the complexity between the agent's task environment, and its sensory, motor, and control system. For example, a very complicated control system is useless if the agent or the environment are extremely simple (Pfeifer & Scheier, 1999, pp. 455–463). A much too complex control system might "overfit" or get stuck in local minima during learning, thus preventing the generation of robust, truly adaptive behavior. A much too simple control system, relatively to the complexity of the environment, may make impossible the adaptation of the agent to the environmental conditions.

4.11 Grounded internal representation

The action an agent performs at some specific moment should not be determined only externally, based on perceptive input, as in a purely reactive system, but neither be entirely planned ahead, as in some classical AI systems. Action should result emergently from the interaction of current perception with previous action and the internal dynamics of the control system. If the behavior of the agent is adaptive, it means that there is a congruity between the performed actions and the the dynamics of the control system, on one hand, with the structure of the environment, on the other hand. Part of the structure of the environment is thus reflected, implicitly, in the structure of the control system. A representation of some environmental situations may dynamically emerge if the agent is able to distinguish different potentialities for action in these situations, as determined by its goal (Bickhard & Ritchie, 1983; Bickhard, 1993, 1999, 2000).

The meaning used here of the term representation is the ability of the agent to include in his decisional process non-trivial considerations about environmental features not currently accessible to sensorimotor exploration. This is not necessarily related to their identifiability by an external observer as particular states of the control system of the agent.

These representations are not imposed by the user or the designer of the system, but are grounded in the sensorimotor interaction of the agent with the environment. This type of representation has a meaning to the cognitive agent itself, and not only to its human observers, as in classical symbol-based artificial systems.

Memory may be based on the same mechanisms that enable perception and action, as it is in humans, for example (Fuster, 1995). Regularities and structural invariance in sensorimotor patterns may then be implicitly internalized during behavior (Berthouze & Tijsseling, 2001; Robbins, 2002). The memorized part of the sensorimotor structures that were previously experienced interacts with current activation patterns. Common structures and invariants may be implicitly detected by the learning process and reflected in memory, for example through a quasi-hebbian mechanism. Abstraction and categorization may thus be based on the same mechanisms that enable perception, action and learning (Robbins, 2002; Indurkhya, 1992). Categories can emerge as singularities in the continuous flow of sensorimotor data. As such, categories are not static, imposed representations but transient characteristics of the state space. They are activated as the system is involved in it. The category is cued by similar instances, which can be a sensory stimulus, some self-driven dynamic exploration, or a performed action (Berthouze & Tijsseling, 2001). This would ensure the self-organization and continuity of learning and adaptation to environmental changes.

Anticipation, internal simulation, imagery and planning may also be based on associations between past perceptions and actions (Hesslow, 2002; Kosslyn & Thompson, 2000; Jeannerod, 1994, 1999), that reflect the previously experienced structure of the environment.

Sensorimotor simulation may ground even abstract reasoning (Barsalou, 1999; Florian, 2002). Embodied artificial agents may thus eventually develop genuine artificial intelligence.

4.12 Grounded symbolic communication

The meaning of symbols used in communication with other agents (either human or artificial) must be grounded in the sensorimotor interaction with the environment. Otherwise, this meaning cannot be accessible to the agent itself, as it was argued theoretically (Bickhard, 1993). It was also shown experimentally that, in humans, the neural activation triggered by the understanding of a word is similar to the activation exhibited during the experiences of the person with the significant of that word (Damasio, 1990; Pulvermuller, 1999; Martin et al., 2000).

Based on their experiments regarding the evolution of language in a population of artificial agents, Steels et al. (2000) have established a set of principles regarding the self-organization of symbolic communication. Agents must be able to engage in coordinated interactions, i.e. to have shared goals and a willingness to cooperate. Agents must have parallel non-verbal ways to achieve the goals of verbal interactions, for example by pointing, gaze following, grasping, etc. This implies that the group of agents involved in communication should share sensorimotor access to the same environment. Agents must have ways to conceptualize reality and to form these conceptualizations, constrained by their embodiment and history and the ontology underlying the emerging lexicon. The concept formation processes of the agents must be based on similar (not necessarily identical) embodiment and result in similar although not necessarily equal conceptual repertoires. The conceptualization for a particular situation must be constrained to be similar so that the agents have a reasonable chance at guessing the conceptualization that a speaker may have used. Agents must have ways to recognize word forms and reproduce them. Also, agents must have the ability to discover and use the strongest associations (between words and meanings) in the group. There must be sufficient group stability to enable a sufficient set of encounters between agents with similar lexicons. Initial group size should not be too large so that there are enough encounters between the same individuals. There must be sufficient environmental stability, in order to have conceptual stability.

4.13 Interdependencies between the principles

There are many interdependencies between the principles stated above. For example, sensorimotor coordination is tightly connected with the embodiment and the situatedness of the agent. Interesting kinds of sensorimotor coordination require a design of the sensorimotor capabilities of the agent that respects the redundancy principle. If the agent is to be autonomous, it has to learn on its own about the environment, and thus to be capable of self-organized learning. Self-organization and emergence is dependent on sensorimotor coordination and ecological balance, and is guided by the agent's goals. A design based on parallel, loosely coupled processes may ensure emergence and may also respect the principles of cheap design and of redundancy. Exploitation of the constrains of the ecological niche may also lead to cheap designs (Pfeifer & Scheier, 1999, pp. 318–319).

5 Design issues

Many issues have to be considered in the design of artificial intelligent agents. We will discuss some of them from the perspective of the design of an agent with the purpose of studying theoretical principles of artificial intelligence.

As we have seen, an agent must be embodied and situated in an environment. The choice of a real (physical) environment versus a simulated one is an important issue.

There are important advantages of using a simulated world, including a simulation of the agent's body. Since most of the hardware considerations may be omitted, there is more time to focus on the conceptual issues. It is much simpler to modify the body of a simulated agent than to modify a preexisting robot: it may require changing a few lines of code, versus many hours of engineering work. A simulated agent may be much cheaper to code, in comparison with the cost of a real robot. In simulation, one does not have to worry about charging the batteries. Common real robots have an autonomy of just several hours, when running on batteries. Simulated robots do not wear off, thus imposing recurrent costs on the experiment, neither break, which may result in unwanted interruptions of the experiments. Simulation of some simple environments, like in navigation experiments, may also be faster than real time. This makes simulation preferable for evolutive methods, where the behavior of generations of agents in the environment has to be tracked for long periods of time.

However, there are also disadvantages of simulation. It is hard to simulate the dynamics of a physical robot and of an environment realistically, especially if the simulated agents have many degrees of freedom. In the real world, the dynamics is simply given by the laws of physics. A simulated environment is always simpler than the real world, with its infinite richness. This simplification is based on the designer's perspective of what features of the environment are important and what are negligible. On one hand, this limits the possible ontologies that the agent may develop. On the other hand, it may limit the capability of the agent to deal with the complexity of the real world.

The tasks of the agent must be defined, in conjunction with the characteristics of its body, sensors and effectors, and the environmental niche. The particular tasks that are chosen depend on the subject of the research.

Their choice should be justified by the stated hypotheses and purposes of the experiment. Many agent studies draw their inspiration regarding the choice of tasks from biological examples. Navigation and related tasks (such as obstacle avoidance, light seeking) and interaction with objects (such as sorting, collecting) are thus common tasks found in the research literature.

The design of the agent's body is another nontrivial issue, constrained by the proposed task, the purpose of the experiment, availability of parts and of necessary budget, availability of qualified personnel for the construction of the agent, and the deadlines established for the project. A component may serve multiple purposes. An arm, for example, may be used for object manipulation, for maintaining balance in walking, for crawling, for protection and attack, and for communication. Possible sensors for physical robots may be cameras (mobile or not, color or grayscale), touch sensors, sonars, infrared sensors, odometers, accelerometers, laser scanners, magnetic compasses, global positioning systems. The redundancy principle (Section 4.9) should be respected in the choice and the positioning of the sensors. Current technologies do not offer many convenient choices for motor systems. Many common physical robots have electrical motors, which power wheel-based locomotion systems, and eventually primitive arms or grabbers. Muscle wires may be used in small robots. More complex motor systems, based on novel synthetic active materials or with many degrees of freedom, may be quite expensive.

A widely used platform for experiments in cognitive robotics, especially for navigation tasks, is Khepera, a miniature mobile robot developed at EPFL, Switzerland and currently produced by K-Team[11]. Khepera has a circular shape with a diameter of 55 mm, a height of 30 mm, and a weight of 70g. Its small size implies that experiments can be performed in regular offices, on a table top. The robot is supported by two wheels and two small teflon balls. The wheels are driven by extremely accurate stepper motors under PID control and can move both forward or backward. The robot is provided with eight infra-red proximity sensors. Six sensors are positioned on the front of the robot, the remaining two on the back. Optional modules ("turrets") can be attached, with cameras or a gripper that can manipulate, in a rather simple fashion, small objects. A Motorola 68331 controller with 256 Kbytes RAM and 512 Kbytes ROM manages all the input-output routines and can communicate via a serial port with a host computer. Khepera may be attached to a host computer by means of a lightweight aerial cable and specially designed rotating contacts, or may operate autonomously, with the controlling program uploaded on the onboard memory. Several simulation programs were created for this platform, for example Webots[12].

The control architecture of the agent a determinant factor for the per-

[11] http://www.k-team.com

[12] http://www.cyberbotics.com/products/webots/

formance of the agent and the most important research issue. Several design methods for the control systems will be presented next (Section 7). Neural networks are a preferred approach to control systems for autonomous agents. They are robust, fault and noise tolerant. They are well adapted to learning and generalization. They respect the principle of parallel, loosely coupled processes and, because of being composed of many simple interacting elements, can display emergence. Because of their many free parameters, especially the weights, they incorporate a sufficient amount of redundancy for adapting to novel situations. They are biologically-inspired, thus facilitating the implementation of architectures inspired by real brains. Spiking neural networks (Maas & Bishop, 1999; Gerstner & Kistler, 2002; Rieke, Warland, Steveninck, & Bialek, 1996) are the type of networks that respect the most closely biological plausibility, in the conditions of the simplifications necessary for the possibility of computational implementations. It was shown that their computational capabilities are more powerful than those of classical, continuous-valued neural networks (Maas, 1997a, 1997b). Their intrinsic temporal character suggest them as suitable for composing control systems for autonomous agents, where real-time performance is needed.

6 Evaluation and analysis

There is no generally accepted way of evaluating intelligent agents, and cognitive science models in general. However, an autonomous agent research project should also include systematic evaluation and analysis of the resulted artificial agent. A simple conclusion regarding the success (or the lack of it) of the agent in the performance of the task, although important, especially in pilot or exploratory studies, is not always sufficient for understanding the scientific relevance of the study.

The evaluation is dependent on the purpose of the project: building a robot for a particular applicative task, studying general principles of intelligence, or modelling certain aspects of biological agents. In many cases, evaluation would include: an assessment of the performance of the desired task; a comparison with biological agents, where possible; compliance with the design principles; an assessment of the heuristic value of the experiment; and a comparison with other approaches.

The most obvious and common evaluation method is the observation of behavior. It may be just a qualitative assessment; in this case, one should not ignore that the interpretation of observed behavior by a human, as well as its segmentation, are highly subjective and dependent on the human perspective and ontologies, which may be very different from the agent's. Some parameters of the agent behavior (like heading directions, distance travelled, various characteristics of movement and action—angles, forces, and so on) may be recorded systematically, and then analyzed. This analysis may be statistical, or in terms of dynamical and mathematical models (e.g. Lerman, Galstyan, Martinoli, & Ijspeert, 2001).

Systematical analysis may also be carried by varying the characteristics of the environment or of the agent's morphology and sensorimotor capabilities (e.g. Bongard & Pfeifer, 2002).

When the control system is based on a neural network, it can be also analyzed from various perspectives, such as dynamical systems theory (Strogatz, 1994; e.g. Beer, 1995), the theory of far-from-equilibrium systems (Haken, 1989), statistical learning theory (Vapnik, 1998), information theory (Shannon & Weaver, 1949; e.g. Rieke et al., 1996; Lungarella & Pfeifer, 2001b, 2001a; Tononi, Sporns, & Edelman, 1994, 1996, 1999). A systematic method for the localization of function in neural networks has been recently proposed (Segev, Aharonov, Meilijson, & Ruppin, 2002; Aharonov, Segev, Meilijson, & Ruppin, 2003).

7 Approaches in autonomous agent research

As we have repeatedly argued throughout this paper, there are many problems with the classical symbolic, modular approaches to the design of systems for autonomous agent control. We will present next several approaches that respect the design principles stated above.

7.1 The subsumption architecture

The subsumption, or behavior-based architecture was introduced in the 80's by Rodney Brooks, currently director of the MIT Artificial Intelligence laboratory, as an engineering solution for the problems of the classical robotic systems (Brooks, 1986; Arkin, 1998; Pfeifer & Scheier, 1999, chap.7). Brooks' intention was to create a methodology that would make it easy to design robots that pursue multiple goals and respond to multiple sensors, perform robustly, and are incrementally extendable.

This architecture was conceived to reflect aspects of natural evolution, such as the idea of having layers that need not be changed once they have been created. It respects the principles of sensorimotor coordination and of the parallel, loosely coupled processes, stated above. Having relatively direct couplings from sensors to actuators leads to good real time performance, because it lacks the time-consuming modelling operations and planning processes from classical systems.

Subsumption is a method of decomposing the control architecture of an agent into a set of task-achieving behaviors or competencies. The classical approach to building control architectures for robots was functional decomposition: information from different sensory systems is integrated in a central representation; a model of the environment is then built or updated; on the basis of the model, an action is planned and executed. In contrast to this approach, the subsumption architecture is built by incrementally adding task-achieving behaviors on top of each other. Implementations of such behaviors are called layers. Higher-level layers (e.g. exploration of the environment) are built and rely on lower-level ones (like object avoidance). Higher layers can subsume lower ones. Instead of having a single sequence of information flow, from perception to model to action, there are multiple paths, the layers, that are active in parallel. Each of these tasks is concerned with only a small subtask of the robot's overall task, such as avoiding walls, circling around targets, or moving to a charging station. Each of these layers can function relatively independently. They do not have to await instructions or results produced by other layers. Thus control is not hierarchical. The subsumption approach realizes direct couplings between sensors and effectors, with only limited internal processing.

The starting point of this architecture is defining levels of competence. A level of competence is the informal specification of a class of desired behaviors that the robot should be able to perform. Each level of competence is implemented as a layer. The layers can be built incrementally, which leads to designs in which new competencies can be added to the already existing and functioning control system (for example, layer 0 for obstacle avoidance, level 1 for exploration, level 2 for collecting objects). Once each layer has been built and debugged, it never has to be changed again. Incremental extendibility is an important factor that contributed for the popularity of this architecture. At each level, there are sensory inputs and motor outputs. Higher levels, like lower ones, can directly interact with the environment, without the need to go through lower levels.

Each layer consists of a set of modules that asynchronously send messages to each other over connecting wires. Each module is an augmented finite state machine. Input to modules can be suppressed and outputs can be inhibited by wires from other modules. Through this mechanism, higher-level layers can subsume lower-level ones, hence the name of the architecture.

Examples of robots that were built using the subsumption architecture are Myrmix (a wheeled robot that find food items in a simple environment and "eats" them), Ghenghis (a hexapod walking robot). Cog[13] is a humanoid robot that was supposed to serve as a platform that would show that higher level, human-like intelligence can emerge in a system based on the subsumption design architecture, i.e. from many, relatively independent processes, based on sensorimotor couplings with relatively little internal processing. However, it appeared that the original architecture had to be extended to include learning, and the principle of not modifying already implemented layers could not be respected (Pfeifer & Scheier, 1999, chap.7).

Until now, systems based on this architecture failed to display extremely interesting intelligent behavior. However, the architecture leads to robust

[13] http://www.ai.mit.edu/projects/humanoid-robotics-group/cog/cog.html

systems, that are useful for a wide range of applications, such as the robots built by the IRobot company[14].

7.2 Evolutionary methods

Artificial evolution of the control system or the morphology of artificial agents is an interesting alternative for their design. By using this methodology, the biases of the human designer are kept to a minimum, thus allowing the possibility of exploring automatically regions of design space that conventional design approaches are often constrained to ignore. In most of the experiments conducted with artificial evolution one can observe the emergence of behavior exploiting sensorimotor coordination to solve difficult tasks. In many cases the evolved solutions are much simpler than those that can be obtained through explicit design.

The analysis of evolved agents and the identification of how they exploit the interaction with the environment is often very difficult and requires significant effort, but is generally much simpler than the analysis of natural organisms because the former are much more simple and can be manipulated much more freely than the latter. Such analysis may allow the identification of new explanatory hypotheses that may produce new models of adaptive behavior and cognition. Evolutionary experiments with autonomous agents also allow a better understanding of principles like sensorimotor coordination, the importance of online learning, and the advantages that arise from the interaction between evolution and lifetime adaptation (Nolfi & Floreano, 2002, 2000; Meyer, 1998; Pfeifer & Scheier, 1999, chap. 8).

The experiments that use evolutionary methods for agent design can be included in two research domains, artificial life and evolutionary robotics. Artificial life (alife) is the study of man-made systems that exhibit behaviors characteristic of natural living systems. It complements the traditional biological sciences by attempting to synthesize life-like behaviors within computers and other artificial media (Langton, 1995; Brooks & Maes, 1994). Evolutionary robotics is the attempt to synthesize robots through evolutionary techniques (Nolfi & Floreano, 2002).

Evolutionary methods usually imply the definition of a fitness function that assesses how well the behavior of the agent comply with its assigned task, and of an encoding scheme that relates the agent's genotype (the information that evolves from generation to generation) to its phenotype (the agent's control architecture or morphology). The fitness is evaluated by letting the agent evolve in the environment for a limited period of time. Evolutionary procedures (such as genetic algorithms, evolutionary strategies or genetic programming) are then used to generate agents with increasing fitness, through mutation and combination of the genotypes.

[14] http://www.irobot.com

In many cases, conducting evolutionary experiments on physical robots can lead to prohibitive problems: the needed running time may be too long, and time also has to be allotted for recharging of batteries, initializing repeatedly the experiment, or repairing defected parts. For some simple environments, robots or behaviors, it was however shown that control systems evolved in simulated environments can be transferred successfully for the control of real robots (Miglino, Lund, & Nolfi, 1995; Jakobi, Husbands, & Harvey, 1995; Jakobi, 1997; Tokura, Ishiguro, Kawai, & Eggenberger, 2001, 2002). For such transfer to be possible, it is important to take into account the fact that identical physical sensors and actuators may actually perform very differently. This problem can be solved by sampling the real world through the sensors and the actuators of the robot. This method, in fact, allows to build a model of a physical individual robot that takes into account the differences between robots of the same type and between different identical components of the same robot. It is also important to account in some way for noise and for other characteristics of the robot and of the environment not included in the simulator (ambient light, slight differences in color and shape of the objects, etc.). This may be realized by introducing in the simulator appropriate profiles of noise and by building a noise-tolerant controller. Too much artificial noise in the simulation may be as deleterious as the lack of it (Jakobi et al., 1995). If a decrease in performance is observed when the system is transferred in the real environment, successful and robust results can be obtained by continuing the evolutionary process in the real environment for a few generations.

Through such methods, agents were evolved that are capable of exploration, obstacle avoidance, wall following, target finding, area cleaning, landmark identification, multiple-legged locomotion (Meyer, 1998), judging the passability of openings relative to their own body size, discriminating between visible parts of themselves and other objects in their environment, predicting and remembering the future location of objects in order to catch them blind, and switching their attention between multiple distal objects (Slocum, Downey, & Beer, 2000). Neural network based control systems can also be evolved to display reinforcement learning-like behavior without modification of the connection strengths (Yamauchi & Beer, 1994b, 1994a; Blynel & Floreano, 2002).

In most cases, only the control systems are evolved, but there also are some experiments involving the evolution of the morphology.

The evolved control systems are usually neural networks that range from simple perceptrons to recurrent discrete-time or continuous-time networks. Recurrent connections result in internal state of the networks that may be used as a dynamic memory, and also may lead to oscillations that are useful in locomotion experiments. Designing efficient encoding methods for genotype-phenotype transformations is an active area of research. Boshy and Ruppin (2002, in press) have recently devised an adaptive, self-organizing compressed encoding of the phenotypic synaptic efficacies of the agent's neurocontroller.

Some recent experiments have also studied agents that, besides the phylogenetical evolution, are also able to adapt their control system during their lifetime. It complements evolution by allowing individuals to adapt to environmental changes that take place during the lifetime of the individual or within few generations, and therefore cannot be tracked by evolution. In addition, plastic individuals can adapt to sensory, motor and environmental change that takes place after the evolutionary process. Learning capability can help and guide evolution by channelling the evolutionary process towards promising directions, and it can significantly speed up the synthesis of viable individuals (the so-called Baldwin effect[15]; Turney, 1996; Parisi, Nolfi, & Cecconi, 1992). Learning might also produce more effective behaviors and facilitate the ability to scale-up to problems that involve a larger search space (Nolfi & Floreano, 1999, 2002).

The power of evolutionary methods is limited by the computational time needed for the exploration through evolution of a large search space of possible solutions (Grand, 1998).

7.3 Biologically inspired, engineered models

Many architectures for autonomous agent control are inspired by biological models or by neuroscientific results. Animals are currently the agents that exhibit the greatest degree of autonomy; it is thus a straightforward consideration to have them as inspiration for the design of artificial agents. However, there might be dangers in following this inspiration too closely, outside experiments especially directed towards biological modelling. Biological systems are not designed optimally: through the evolutionary process, solutions were "patched" onto previously working systems. Many vestigial neurological structures, interactions, and side effects may exist in animal brains. Developmental processes needed for the growth and specialization of cells from a single-celled zygote, the supportive mechanisms needed for the nutrition of neurons, and other biological constraints may also lead to side effects in the biological neural architectures. The emulation of these side effects in artificial systems may be a distraction. Moreover, most experiments in neuroscience are still dominated by representational paradigms, in the tradition of classical cognitive science. Autonomous agent research may influence neuroscience experimental paradigms by insisting on the importance of the principle of sensorimotor coordination (Ruppin, 2002), which may lead (circularly), in the long term, to better experimental support for the inspiration needed to its own development.

Biologically inspired ideas pervade, to various degrees and at various

[15] http://www.cs.bath.ac.uk/~jjb/web/baldwin.html

levels, most of the work in simulation of adaptive behavior with artificial agents (animats) (Hallam, Floreano, Hallam, Hayes, & Meyer, 2002; Meyer, Berthoz, Floreano, Roitblat, & Wilson, 2000; Pfeifer, Blumberg, Meyer, & Wilson, 1998; Maes, Mataric, Meyer, Pollack, & Wilson, 1996; Cliff, Husbands, Meyer, & Wilson, 1994; Meyer, Roitblat, & Wilson, 1993; Meyer & Wilson, 1991). For example, Banquet, Gaussier, Quoy, Revel, and Burnod (2002) implemented navigational capabilities in a robot controlled by a neural network inspired by several hippocampal subsystems. A context-independent map in the modelled subiculum and entorrhinal cortex encodes essentially the spatial layout of the environment on the basis of a local dominance of ideothetic movement-related information over allothetic (visual) information. A task and temporal context dependent map based on the transition cells in the CA3-CA1 areas allows encoding maps, in higher order structures, as graphs resulting from combination of learned sequences of events. On the basis of these two maps two distinct goal-oriented navigation strategies emerge: one based on a population vector code of the location-action pairs to learn and implement goal reaching; and another one based on linking transition cells together as conditioning chains that will be implemented under the top-down guidance of drives and motivations. Various other biologically inspired models for robot navigation are reviewed by Franz and Mallot (2000).

7.4 Collective behavior, modular robotics

The interaction of a group of agents, even simple ones, may lead to interesting emergent collective behaviors at the group level. Common examples are given by social insects—ants, termites, bees and wasps—and by swarming, flocking, herding, and shoaling phenomena in groups of vertebrates. The abilities of such systems appear to transcend the abilities of the constituent individual agents. In most biological cases studied so far, the robust and capable high level group behavior has been found to be mediated by nothing more than a small set of simple low level interactions between individuals, and between individuals and the environment. The swarm intelligence approach emphasizes distributedness and exploitation of direct (agent-to-agent) or indirect (via the environment) local interactions among relatively simple agents. The main advantages of the application of the swarm approach to the control of a group of robots are three-fold: (1) scalability: the control architecture is kept exactly the same from a few units to thousands of units; (2) flexibility: units can be dynamically added or removed, they can be given the ability to reallocate and redistribute themselves in a self-organized way; (3) robustness: the resulting collective system is robust not only through unit redundancy but also through the unit minimalist design. (Lerman et al., 2001).

In the last few years, the swarm intelligence control principles have been successfully applied to a series of case studies in collective robotics: aggregation and segregation, beacon and odor localization, collaborative mapping, collaborative transportation, work division and task allocation, flocking and foraging. All these tasks have been performed using groups of simple, autonomous robots or embodied simulated agents, exploiting local communication forms among teammates (implicit, through the environment, or explicit, wireless communication), and fully distributed control.

For example, Beckers, Holland, and Deneubourg (1994) designed an experiment where robots are equipped with a forward-facing C-shaped gripper which is able to collect small pucks from the environment, two infrared sensors for obstacle avoidance, and a microswitch which is activated by the gripper when a certain number of pucks are pushed. The robots have only three behaviors, and only one is active at any time. When no sensor is activated, a robot executes the default behavior of moving in a straight line until an obstacle is detected or until the microswitch is activated (pucks are not detected as obstacles). On detecting an obstacle, the robot executes the obstacle avoidance behavior of turning on the spot away from the obstacle and through a random angle; the default behavior then takes over again, and the robot moves in a straight line in the new direction. If the robot is pushing pucks when it encounters the obstacle, the pucks will be retained by the gripper throughout the turn. When the gripper pushes three or more pucks, the microswitch is activated; this triggers the puck-dropping behavior, which consists of backing up by reversing both motors for 1 second (releasing the pucks from the gripper), and then executing a turn through a random angle, after which the robot returns to its default behavior and moves forwards in a straight line. The obstacle avoidance behavior has priority over the puck-dropping behavior. There is no communication between the robots, all they do is performing these simple three behaviors. However, the result of the experiment (involving five robots) is that the pucks, initially dispersed randomly in the environment, are gathered in clusters.

Self-organizing multiagent societies may have interesting applications in services (such as vacuuming and cleaning), industry (assembly) and defense (for surveillance and transport). The "smart dust" concept implies sprinkling thousands of tiny wireless sensors on a battlefield to monitor enemy movements without alerting the enemy to their presence. By self-organizing into a sensor network, smart dust would filter raw data for relevance before relaying only the important findings to central command. The idea was launched by the Pentagon in 1999 and has recently reached prototype stage[16]. Brooks and Flynn (1989) has proposed the used of robotic swarms for space exploration.

Modular reconfigurable robotics is a related approach to building robots for various complex tasks. Robots are built out of a number of identical

[16] http://www.eet.com/at/news/OEG20030128S0028

simple modules. Each module contains a processing unit, a motor, sensors and the ability to attach to other modules. One module can't do much by itself, but when many modules are connected together, the result may be a system capable of complex behaviors. A modular robot can even reconfigure itself—change its shape by moving its modules around—to meet the demands of different tasks or different working environments. For example, the PolyBot developed at the Palo Alto Research Center is capable to reconfigure itself for snake-like or spider-like locomotion (Yim, Duff, & Roufas, 2000). Self-reconfigurable robots have several advantages over traditional fixed-shape robots: (1) The modules can connect in many ways making it possible for a single robotic system to solve a range of tasks. This is useful in scenarios where is undesirable to build a special purpose robot for each task. The robot may even decompose itself in several smaller ones. (2) Self-reconfigurable robots can adapt to the environment and change shape as needed. (3) Since the robot is built out of many independent modules it can be robust to module failures. Defect modules can be ejected from the system and the robot may still perform its task. (4) Modules can be mass-produced and therefore the cost may be kept low (Støy, Shen, & Will, 2002). These robots may have applications in rescue scenarios from collapsed buildings, where they may enter the rubble with snake-like locomotion and then reconfigure to support the weight of the rubble collapsed on the buried people. Their ability to serve as many tools at once, versatility and robustness recommend them for space applications, saving weight and being able to packing into compressed forms. They may also have various military applications. However, there currently exist a number of both hardware and software research challenges that have to be overcome in order for modular robots to be able to perform in such applications.

8 Embodied agents as far-from-equilibrium systems

The self-organized formation of structures in far-from-equilibrium systems was observed in different branches of physics, chemistry, mechanical engineering and biology (Haken, 1989; Cross & Hohenberg, 1993; Prigogine & Stengers, 1984). This type of emergence may be eventually exploited by novel computational paradigms.

Embodied autonomous agents represent an attractive framework for the study of the computational properties of far-from-equilibrium systems. On one hand, the self-organizational characteristics of these systems recommend them as support for the emergence of interesting adaptational and cognitive properties in autonomous agents (Smithers, 1995, p. 153). On the other hand, an autonomous agent, through continuous dynamic interaction with the environment, provides a non-trivial sustained input to the system that may keep it in a non-equilibrium state. Internal self-driven dynamics, such as threshold phenomena triggered by coincidence of random self-excitation, may also contribute to avoidance of stable or trivial states (Berthouze & Tijsseling, 2001). Spiking neural networks seem to be an ideal support for autonomous agent non-equilibrium control systems, because of their intrinsic temporal characteristics, computational capabilities and biological plausibility (Maas & Bishop, 1999; Gerstner & Kistler, 2002; Rieke et al., 1996; Maas, 1997a, 1997b). They might sustain complex, chaotic dynamics (Banerjee, 2001b, 2001a). Considerations from the stochastic dynamical systems theory (Freeman, Kozma, & Werbos, 2001) may have to be taken into account.

9 Conclusion

Autonomous intelligent agent research is a complex, interdisciplinary domain. As shown above, a particular interest of this research direction is that the obtention of genuine artificial intelligence, with all its possible applications, is dependent on the advances in this research field. We have reviewed here several principles that should guide autonomous agent research. However, it is not easy to reconcile all these principles in a concrete artefact. The complexity of the issues involved prevents the obtention of important advances in this area. But these complex interdependencies are the premises of the emergence of genuine intelligence. Inspirations from the theory of non-equilibrium, stochastic dynamical systems may give a theoretical framework for the study of self-organization in autonomous agent control.

References

Aharonov, R., Segev, L., Meilijson, I., & Ruppin, E. (2003). Localisation of function via lesion analysis. *Neural Computation*, *15*. (Available from: http://www.cs.tau.ac.il/~ruppin/nc02.ps.gz)

Alissandrakis, A., Nehaniv, C. L., & Dautenhahn, K. (2001). Through the looking-glass with ALICE—trying to imitate using correspondences. In C. Balkenius, J. Zlatev, H. Kozima, K. Dautenhahn, & C. Breazeal (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund, Sweden. Lund University Cognitive Studies, 85.* Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/epigenetic-robotics/Papers/Alissandrakis.pdf)

Anderson, J. R. (1993). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Andry, P., Gaussier, P., & Nadel, J. (2002). From visuo-motor development to low-level imitation. In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland. Lund University Cognitive Studies, 94* (pp. 7–15). Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/ftp/pub/LUCS_Studies/LUCS94/Andry.pdf)

Arkin, R. C. (1998). *Behavior-based robotics.* Cambridge, MA: MIT Press.

Balkenius, C., Zlatev, J., Kozima, H., Dautenhahn, K., & Breazeal, C. (Eds.). (2001). *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund, Sweden. Lund University Cognitive Studies, 85.* Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/Abstracts/LUCS_Studies/LUCS85.html)

Banerjee, A. (2001a). On the phase-space dynamics of systems of spiking neurons. II: Formal analysis. *Neural Computation*, *13*, 195–225. (Available from: http://www.bcs.rochester.edu/people/arunavab/papers/nc-neuraldynamics2.pdf)

Banerjee, A. (2001b). On the phase-space dynamics of systems of spiking neurons. I: Model and experiments. *Neural Computation*, *13*, 161–193. (Available from: http://www.bcs.rochester.edu/people/arunavab/papers/nc-neuraldynamics1.pdf)

Banquet, J. P., Gaussier, P., Quoy, M., Revel, A., & Burnod, Y. (2002). Cortico-hippocampal maps and navigation strategies in robots and rodents. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior* (pp. 141–150). Cambridge, MA: MIT Press. (Available from: http://www-etis.ensea.fr/~neurocyber/hs3_sab02_banquet_gaussier.ps)

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660. (Available from: http://www.bbsonline.org/Preprints/OldArchive/bbs.barsalou.html)

Beckers, R., Holland, O., & Deneubourg, J. (1994). From local actions to global tasks: Stigmergy and collective robotics. In R. Brooks & P. Maes (Eds.), *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (pp. 181–189). Cambridge, MA: MIT Press.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, *72*, 173–215. (Available from: http://vorlon.ces.cwru.edu/~beer/Papers/AIJ95.pdf)

Berthouze, L., & Tijsseling, A. (2001). Embodiment is meaningless without adequate neural dynamics. In R. Pfeifer, G. Westermann, C. Breazeal, Y. Demiris, M. Lungarella, R. Nunez, & L. Smith (Eds.), *Proceedings of the Workshop on Developmental Embodied Cognition, Edinburgh* (pp. 21–25). Edinburgh, Scotland. (Available from: http://www.cogsci.ed.ac.uk/~deco/posters/berthouze.pdf)

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, 285–333. (Available from: http://www.lehigh.edu/~mhb0/repconpage.html)

Bickhard, M. H. (1999). The dynamics of representation. In B. Hayes, C. Hooker, R. Heath, & A. Heathcote (Eds.), *Proceedings of the Fourth Australian Cognitive Science Conference.* Newcastle, Australia: University of Newcastle.

Bickhard, M. H. (2000). Dynamic representing and representational dynamics. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bickhard, M. H., & Ritchie, D. M. (1983). *On the nature of representation.* New York, NY: Praeger.

Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Blake, A., & Yuille, A. (Eds.). (1992). *Active vision.* Cambridge, MA: MIT Press.

Blynel, J., & Floreano, D. (2002). Levels of dynamics and adaptive behavior in evolutionary neural controllers. 272–281. (Available from: :http://asl.epfl.ch/aslInternalWeb/ASL/publications/uploadedFiles/blynel_sab02.pdf)

Bongard, J. C., & Pfeifer, R. (2002). A method for isolating morphological effects on evolved behaviour. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior* (pp. 305–311). Cambridge, MA: MIT Press. (Available from: http://www.ifi.unizh.ch/ailab/people/bongard/papers/bongardPfeiferSAB2002.ps.gz)

Boshy, S., & Ruppin, E. (2002). Small is beautiful: Near minimal evolutionary controllers obtained with self-organizing compressed encoding. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior* (pp. 345–346). Cambridge, MA: MIT Press.

Boshy, S., & Ruppin, E. (in press). Evolution of near minimal agents with a self-organized compressed encoding. *Artificial Life*. (Available from: http://www.cs.tau.ac.il/~ruppin/SOCEjurnal.ps.gz)

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology.* Cambridge, MA: MIT Press.

Brooks, R., & Maes, P. (Eds.). (1994). *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems.* 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. (Available from: http://www.ai.mit.edu/people/brooks/papers/AIM-864.pdf)

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*, 3–15. (Available from: http://www.ai.mit.edu/people/brooks/papers/elephants.pdf)

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, *47*, 139–159. (Available from: http://www.ai.mit.edu/people/brooks/papers/representation.pdf)

Brooks, R. A. (1995). Intelligence without reason. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 25–81). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brooks, R. A., & Flynn, A. M. (1989). Fast, cheap and out of control: A robot invasion of the solar system. *Journal of the British Interplanetary Society*, *42*, 478–485. (Available from: http://www.ai.mit.edu/people/brooks/papers/fast-cheap.pdf)

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, *20*, 553–557. (Available from: http://vorlon.ces.cwru.edu/~beer/Papers/TINS.pdf)

Churchland, P., & Sejnowski, T. J. (1994). *The computational brain.* Cambridge, MA: MIT Press.

Clancey, W. J. (1995). A boy scout, Toto, and a bird: How situated cognition is different from situated robotics. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 227–236). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, A., & Thornton, C. (1997). Trading spaces: Computation, representation and the limits of uninformed learning. *Behavioral and Brain Sciences*, *20*, 57–92. (Available from: http://www.bbsonline.org/Preprints/OldArchive/bbs.clark.html)

Cliff, D., Husbands, P., Meyer, J.-A., & Wilson, S. W. (Eds.). (1994). *From animals to animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Cross, M. C., & Hohenberg, P. C. (1993). Pattern formation outside of equilibrium. *Reviews of Modern Physics*, *65*.

Damasio, A. (1990). Category-related recognition defects as as a clue to the neural substrates of knowledge. *Trends in Neurosciences*, *13*, 95–98.

Epstein, S. L. (1994). For the right reasons: The FORR architecture for learning in a skill domain. *Cognitive Science*, *18*, 479–511.

Feynman, R. P. (1965 / 1992). *The character of physical law.* London, UK: Penguin Books.

Florian, R. V. (2002). Why it is important to build robots capable of doing science. In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland. Lund University Cognitive Studies, 94* (pp. 27–34). Lund, Sweden: Lund University. (Available from: http://www.coneural.org/florian/papers/robotic_science_02.php)

Fodor, J. (1975). *The language of thought.* New York, NY: Crowell.

Franceschini, N., Pichon, J. M., & Blanes, C. (1992). From insect vision to robot vision. *Philosophical Transactions of the Royal Society, London B*, *337*, 283–294.

Franklin, S., & Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages.* Springer Verlag. (Available from: http://www.msci.memphis.edu/~franklin/AgentProg.html)

Franz, M. O., & Mallot, H. A. (2000). Biomimetic robot navigation. *Robotics and Autonomous Systems*, *30*, 133–153. (Available from: http://www.uni-tuebingen.de/cog/personpages/ham/publication/epapers/roboticsautonomoussystems00.pdf)

Freeman, W. J., Kozma, R., & Werbos, P. J. (2001). Biocomplexity: adaptive behavior in complex stochastic dynamical systems. *BioSystems*, *59*, 109–123. (Available from: http://www.msci.memphis.edu/~kozmar/biosys.pdf)

Fuster, J. M. (1995). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate.* Cambridge, MA: MIT Press.

Gerstner, W., & Kistler, W. M. (2002). *Spiking neuron models.* Cambridge, UK: Cambridge University Press. (Available from: http://diwww.epfl.ch/~gerstner/BUCH.html)

Grand, S. (1998). Battling with GA-Joe. *IEEE Intelligent Systems*, *13*, 18–20. (Available from: http://www.cyberlife-research.com/articles/ieee/ieee3.htm)

Haken, H. (1989). Synergetics: an overview. *Reports of Progess in Physics*, *52*, 515–533.

Hallam, B., Floreano, D., Hallam, J., Hayes, G., & Meyer, J.-A. (Eds.). (2002). *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Hallam, J. (1995). Autonomous robots: A question of design? In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 217–226). Hillsdale, NJ: Lawrence Erlbaum Associates.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335-346. (Available from: http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.sgproblem.html)

Held, R., & Hein, A. (1958). Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual Motor Skills*, *8*, 87–90.

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, *6*, 242–247.

Iantovics, B., & Dumitrescu, D. (in press). *Agenţi artificiali inteligenţi [Intelligent artificial agents].* Romania.

Indurkhya, B. (1992). *Metaphor and cognition: An interactionist approach.* Dordrecht, the Nederlands: Kluwer Academic Publishers.

Jakobi, N. (1997). Evolutionary robotics and the radical envelope of noise hypothesis. *Adaptive Behavior*, *6*, 131–174. (Available from: ftp://ftp.cogs.susx.ac.uk/pub/reports/csrp/csrp457.ps.Z)

Jakobi, N., Husbands, P., & Harvey, I. (1995). Noise and the reality gap: The use of simulation in evolutionary robotics. In F. Moran, A. Moreno, J. Merelo, & P. Chacon (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life. Lecture Notes in Artificial Intelligence, 929* (pp. 704–720). (Available from: http://citeseer.nj.nec.com/jakobi95noise.html)

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*, 187–245. (Available from: http://www.bbsonline.org/Preprints/OldArchive/bbs.jeannerod.html)

Jeannerod, M. (1999). *The cognitive neuroscience of action.* Blackwell.

Kolesnik, M., & Streich, H. (2002). Visual orientation and motion control of MAKRO - adaptation to the sewer environment. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Kosslyn, S. M., & Thompson, W. L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences, 2nd edition* (pp. 975–985). Cambridge, MA: MIT Press.

Kozima, H., Nagakawa, C., & Yano, H. (2002). Emergence of imitation mediated by objects. In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, UK. Lund University Cognitive Studies, 94* (pp. 59–61). Lund, Sweden. (Available from: http://www.lucs.lu.se/ftp/pub/LUCS_Studies/LUCS94/Kozima.pdf)

Lakoff, G., & Nunez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* Basic Books.

Lambrinos, D., Maris, M., Kobayashi, H., Lebhart, T., Pfeifer, R., & Welmer, R. (1997). An autonomous agent navigating with a polarized light compass. *Adaptive Behavior*, *6*, 175–206.

Langton, C. G. (Ed.). (1995). *Artificial life: An overview.* Cambridge, MA: MIT Press.

Lehky, S. R., & Sejnowski, T. J. (1988). Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature*, *333*, 452–454.

Lerman, K., Galstyan, A., Martinoli, A., & Ijspeert, A. J. (2001). A macroscopic analytical model of collaboration in distributed robotic systems. *Artificial Life*, *7*, 375–393. (Available from: http://www.isi.edu/~lerman/papers/lerman-alife.pdf)

Lungarella, M., & Berthouze, L. (2002). Adaptivity through physical immaturity. In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland. Lund University Cognitive Studies, 94.* Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/ftp/pub/LUCS_Studies/LUCS94/Lungarella.pdf)

Lungarella, M., & Pfeifer, R. (2001a). Robots as cognitive tools: Information theoretic analysis of sensory(-motor) data. In R. Pfeifer, G. Westermann, C. Breazeal, Y. Demiris, M. Lungarella, R. Nunez, & L. Smith (Eds.), *Proceedings of the Workshop on Developmental Embodied Cognition, Edinburgh* (pp. 11–12). Edinburgh, Scotland. (Available from: http://www.cogsci.ed.ac.uk/~deco/invited/max.pdf)

Lungarella, M., & Pfeifer, R. (2001b). Robots as cognitive tools: Information theoretic analysis of sensory-motor data. In R. Pfeifer, Y. Kuniyoshi, O. Sporns, G. Metta, G. Sandini, & R. Nunez (Eds.), *Proceedings of the Workshop on Emergence and Development of Embodied Cognition, Beijing* (pp. 18–25). Beijing, China. (Available from: http://www.ifi.unizh.ch/ailab/people/lunga/Conferences/EDEC2/MaxLungarella.pdf)

Maas, W. (1997a). Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, *10*, 1659–1671. (Available from: http://www.cis.tugraz.at/igi/maass/psfiles/85a.pdf)

Maas, W. (1997b). Noisy spiking neurons with temporal coding have more computational power than sigmoidal neurons. In M. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9, pp. 211–217). Cambridge, MA: MIT Press. (Available from: http://www.cis.tugraz.at/igi/maass/psfiles/90.pdf)

Maas, W., & Bishop, C. M. (Eds.). (1999). *Pulsed neural networks.* Cambridge, MA: MIT Press.

Maes, P. (1995). Artificial life meets entertainment: Life like autonomous agents. *Communications of the ACM*, *38*, 108–114.

Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., & Wilson, S. W. (Eds.). (1996). *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences, 2nd edition* (p. 1023-1036). Cambridge, MA: MIT Press.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding.* Boston, MA: Shambala.

McCarthy, J. (1984). Some expert systems need common sense. *Annals of the New York Academy of Sciences*, *426*. (Available from: http://www-formal.stanford.edu/jmc/someneed.html)

Meyer, J.-A. (1998). Evolutionary approaches to neural control in mobile robots. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics.* (Available from: http://citeseer.nj.nec.com/meyer98evolutionary.html)

Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H. L., & Wilson, S. W. (Eds.). (2000). *From animals to animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Meyer, J.-A., Roitblat, H. L., & Wilson, S. W. (Eds.). (1993). *From animals to animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Meyer, J.-A., & Wilson, S. W. (Eds.). (1991). *From animals to animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Miglino, O., Lund, H. H., & Nolfi, S. (1995). Evolving mobile robots in simulated and real environments. *Artificial Life*, *2*, 417–434. (Available from: http://gral.ip.rm.cnr.it/nolfi/papers/miglino.sim-real.pdf)

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nolfi, S., & Floreano, D. (1999). Learning and evolution. *Autonomous robots*, *7*, 89–113. (Available from: http://gral.ip.rm.cnr.it/nolfi/papers/nolfi.evo-learn.pdf)

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines.* Cambridge, MA: MIT Press.

Nolfi, S., & Floreano, D. (2002). Synthesis of autonomous robots through evolution. *Trends in Cognitive Science*, *6*, 31–37.

Oka, N., Morikawa, K., Komatsu, T., Suzuki, K., Hiraki, K., Ueda, K., & Omori, T. (2001). Embodiment without a physical body. In R. Pfeifer, G. Westermann, C. Breazeal, Y. Demiris, M. Lungarella, R. Nunez, & L. Smith (Eds.), *Proceedings of the Workshop on Developmental Embodied Cognition, Edinburgh.* Edinburgh, Scotland. (Available from: http://www.cogsci.ed.ac.uk/~deco/posters/oka.pdf)

O'Regan, J., & Noe, A. (in press). A sensorimotor account of vision and visual consciousness. *Behavioural and Brain Sciences*, *24*, 5. (Available from: http://www.bbsonline.org/Preprints/ORegan/)

Parisi, D., Nolfi, S., & Cecconi, F. (1992). *Learning, behavior and evolution.* Cambridge, MA: MIT Press. (Available from: http://gral.ip.rm.cnr.it/nolfi/papers/parisi.lbe.pdf)

Pfeifer, R., Blumberg, B., Meyer, J.-A., & Wilson, S. W. (Eds.). (1998). *From animals to animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Pfeifer, R., Kuniyoshi, Y., Sporns, O., Metta, G., Sandini, G., & Nunez, R. (Eds.). (2001). *Proceedings of the Workshop on Emergence and Development of Embodied Cognition, Beijing.* Beijing, China. (Available from: http://www.ifi.unizh.ch/ailab/people/lunga/Conferences/EDEC2/edec-proceedings.pdf)

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence.* Cambridge, MA: MIT Press.

Pfeifer, R., Westermann, G., Breazeal, C., Demiris, Y., Lungarella, M., Nunez, R., & Smith, L. (Eds.). (2001). *Proceedings of the Workshop on Developmental Embodied Cognition, Edinburgh.* Edinburgh, Scotland. (Available from: http://www.cogsci.ed.ac.uk/~deco/deco-proceedings.pdf)

Popper, K. R. (1959). *The logic of scientific discovery.* London, UK: Hutchinson.

Prigogine, I., & Stengers, I. (1984). *Oredr out of chaos: Man's new dialogue with nature.* Glasgow, UK: Fontana.

Prince, C. G., Demiris, Y., Marom, Y., Kozima, H., & Balkenius, C. (Eds.). (2002). *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland. Lund University Cognitive Studies, 94.* Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/Abstracts/LUCS_Studies/LUCS94.html)

Pulvermuller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336. (Available from: http://www.bbsonline.org/Preprints/OldArchive/bbs.pulvermueller.html)

Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundation of cognitive science. *Behavioral and Brain Sciences*, *3*, 111–132.

Quick, T., Dautenhahn, K., Nehaniv, C., & Roberts, G. (1999). The essence of embodiment: A framework for understanding and exploiting structural coupling between system and environment. *Proceedings of the Third International Conference on Computing Anticipatory Systems, Liege, Belgium. August 9-14, 1999 (CASYS'99).* (Available from: http://homepages.feis.herts.ac.uk/~comqkd/quick_casys99.ps)

Riegler, A. (2002). When is a system embodied? *Cognitive Systems Research*, *3*, 339–348. (Available from: http://pespmc1.vub.ac.be/riegler/papers/riegler02embodiment.pdf)

Rieke, F., Warland, D., Steveninck, R. de Ruyter van, & Bialek, W. (1996). *Spikes: Exploring the neural code.* Cambridge, MA: MIT Press.

Robbins, S. E. (2002). Semantics, experience and time. *Cognitive Systems Research*, *3*, 301–337.

Rosenschein, S. J. (1999). Intelligent agent architecture. In R. A. Wilson & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences.* Cambridge, MA: MIT Press.

Ruppin, E. (2002). Evolutionary embodied agents: A neuroscience perspective. *Nature Reviews Neuroscience*, *3*, 132–142. (Available from: http://www.cs.tau.ac.il/~ruppin/npaper10.ps.gz)

Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Englewood Cliffs, NJ: Prentice Hall.

Segev, L., Aharonov, R., Meilijson, I., & Ruppin, E. (2002). Localisation of function in neurocontrollers. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior* (pp. 161–170). Cambridge, MA: MIT Press.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication.* Chicago: University of Illinois Press.

Shortliffe, E. H. (1976). *Computer-based medical consultations: MYCIN.* New York, NY: American Elsevier.

Simon, H. A., & Kaplan, C. A. (1989). Foundations of cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science* (p. 40). Cambridge, MA: MIT Press.

Sinha, C. (2001). The epigenesis of symbolization. In C. Balkenius, J. Zlatev, H. Kozima, K. Dautenhahn, & C. Breazeal (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund, Sweden. Lund University Cognitive Studies, 85.* Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/epigenetic-robotics/Papers/Sinha.pdf)

Slocum, A. C., Downey, D. C., & Beer, R. D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J.-A. Meyer, A. Berthoz, D. Floreano, H. L. Roitblat, & S. W. Wilson (Eds.), *From animals to animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* (pp. 430–439). Cambridge, MA: MIT Press. (Available from: http://vorlon.ces.cwru.edu/~beer/Papers/SAB2000.pdf)

Smithers, T. (1995). Are autonomous agents information processing systems? In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 123–162). Hillsdale, NJ: Lawrence Erlbaum Associates.

Steels, L. (1995). Building agents out of autonomous behavior systems. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 83–121). Hillsdale, NJ: Lawrence Erlbaum Associates.

Steels, L., & Brooks, R. (Eds.). (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Steels, L., Kaplan, F., McIntyre, A., & Looveren, J. van. (2000). Crucial factors in the origins of word-meaning. In J.-L. Dessalles & L. Ghadakpour (Eds.), *Proceedings of the 3rd Evolution of Language Conference* (pp. 214–217). Paris, France: ENST. (Available from: http://www.csl.sony.fr/downloads/papers/2000/steels-evolang2000.pdf)

Støy, K., Shen, W.-M., & Will, P. (2002). The use of sensors in self-reconfigurable robots. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior* (pp. 48–57). Cambridge, MA: MIT Press.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos with applications to physics, biology, chemistry, and engineering.* Cambridge, MA: Perseus Books.

Terzopoulos, D. (1999). Artificial life for computer graphics. *Communications of the ACM*, *42*, 32–42.

Tokura, S., Ishiguro, A., Kawai, H., & Eggenberger, P. (2001). The effect of neuromodulations on the adaptability of evolved neurocontrollers. In J. Kelemen & P. Sosik (Eds.), *Proceedings of the Sixth European Conference on Artificial Life (ECAL2001). Lecture Notes in Artificial Intelligence 2159* (pp. 292–295). Springer. (Available from: http://www.cmplx.cse.nagoya-u.ac.jp/~tokura/study/paper/Ecal01.ps)

Tokura, S., Ishiguro, A., Kawai, H., & Eggenberger, P. (2002). Analysis of adaptability of evolved neurocontroller with neuromodulations. In M. Gini, W.-M. Shen, & H. Yuasa (Eds.), *Intelligent Autonomous Systems 7* (pp. 341–348). (Available from: http://www.cmplx.cse.nagoya-u.ac.jp/~tokura/study/paper/IAS7.PS.gz)

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences USA*, *91*, 5033–5037. (Available from: http://www.pnas.org/cgi/reprint/91/11/5033.pdf)

Tononi, G., Sporns, O., & Edelman, G. M. (1996). A complexity measure for selective matching of signals by the brain. *Proceedings of the National Academy of Sciences USA*, *93*, 3422–3427. (Available from: http://www.pnas.org/cgi/reprint/93/8/3422.pdf)

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National*

Academy of Sciences USA, *96*, 3257–3262. (Available from: http://www.pnas.org/cgi/reprint/96/6/3257.pdf)

Turney, P. (1996). Myths and legends of the Baldwin effect. *Proceedings of the Workshop on Evolutionary Computing and Machine Learning at the 13th International Conference on Machine Learning (ICML-96), Bari, Italy*, 135–142. (Available from: http://citeseer.nj.nec.com/turney96myths.html)

Vapnik, V. N. (1998). *Statistical learning theory.* New York, NY: John Wiley and Sons.

Varela, F. J. (1995). The re-enchantment of the concrete. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence: Building embodied, situated agents* (pp. 11–22). Hillsdale, NJ: Lawrence Erlbaum Associates.

Varela, F. J., Thompson, E., & Rosch, E. (1992). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.

Webb, B. (1994). Robotic experiments in cricket phonotaxis. In D. Cliff, P. Husbands, J.-A. Meyer, & S. W. Wilson (Eds.), *From animals to animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior* (pp. 45–54). Cambridge, MA: MIT Press.

Webb, B., & Consi, T. R. (Eds.). (2001). *Biorobotics.* Cambridge, MA: MIT Press.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., & Thelen, E. (2001). Artificial intelligence: autonomous mental development by robots and animals. *Science*, *291*, 599–600. (Available from: http://www.cse.msu.edu/dl/SciencePaper.pdf)

Weng, J., & Zhang, Y. (2002). Developmental robots - a new paradigm. In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius (Eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Edinburgh, Scotland. Lund University Cognitive Studies, 94* (pp. 163–174). Lund, Sweden: Lund University. (Available from: http://www.lucs.lu.se/ftp/pub/LUCS_Studies/LUCS94/Weng.pdf)

Yamauchi, B. M., & Beer, R. D. (1994a). Integrating reactive, sequential, and learning behavior using dynamical neural network. In D. Cliff, P. Husbands, J.-A. Meyer, & S. W. Wilson (Eds.), *From animals to animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Yamauchi, B. M., & Beer, R. D. (1994b). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, *2*, 219–246. (Available from: `http://citeseer.nj.nec.com/yamauchi94sequential.html`)

Yim, M., Duff, D., & Roufas, K. (2000). PolyBot: a modular reconfigurable robot. *IEEE International Conference on Robotics and Automation (ICRA).*

Ziemke, T. (2001a). Are robots embodied? In C. Balkenius, J. Zlatev, H. Kozima, K. Dautenhahn, & C. Breazeal (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund, Sweden. Lund University Cognitive Studies, 85.* Lund, Sweden: Lund University. (Available from: `http://www.lucs.lu.se/epigenetic-robotics/Papers/Ziemke.pdf`)

Ziemke, T. (2001b). Disentangling notions of embodiment: Does a robot have a body? In R. Pfeifer, G. Westermann, C. Breazeal, Y. Demiris, M. Lungarella, R. Nunez, & L. Smith (Eds.), *Proceedings of the Workshop on Developmental Embodied Cognition, Edinburgh.* Edinburgh, Scotland. (Available from: `http://www.cogsci.ed.ac.uk/~deco/invited/ziemke.pdf`)

Ziemke, T. (2001c). The construction of 'reality' in the robot: Constructivist perspectives on situated artificial intelligence and adaptive robotics. *Foundations of Science*, *6*, 163–233. (Available from: `http://researchindex.com/ziemke00construction.html`)

Zlatev, J., & Balkenius, C. (2001). Introduction: Why "epigenetic robotics"? In C. Balkenius, J. Zlatev, H. Kozima, K. Dautenhahn, & C. Breazeal (Eds.), *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund, Sweden. Lund University Cognitive Studies, 85.* Lund, Sweden: Lund University. (Available from: `http://www.lucs.lu.se/epigenetic-robotics/Papers/Zlatev.Balkenius.2001.pdf`)