Year: 2005

Interacting with the real world: design principles for intelligent systems

Pfeifer, Rolf ; Gómez, Gabriel

Abstract: The last two decades in the field of artificial intelligence have clearly shown that true intelligence always requires the interaction of an agent with a real physical and social environment. The concept of embodiment that has been introduced to designate the modern approach to designing intelligence has far-reaching implications. Rather than studying computation alone, we must consider the interplay between morphology, materials, brain (control), and the environment. A number of case studies are presented, and it is demonstrated how artificial evolution and morphogenesis can be used to systematically investigate this interplay. Taking these ideas into account requires entirely novel ways of thinking, and often leads to surprising results

DOI: https://doi.org/10.1007/s10015-004-0343-3

Originally published at:
Pfeifer, Rolf; Gómez, Gabriel (2005). Interacting with the real world: design principles for intelligent systems. Artificial Life and Robotics, 9(1):1-6.
DOI: https://doi.org/10.1007/s10015-004-0343-3

Artif Life Robotics (2005) 9:1–6
DOI 10.1007/s10015-004-0343-3

PLENARY TALK

Rolf Pfeifer · Gabriel Gómez

Interacting with the real world: design principles for intelligent systems

Received and accepted: December 13, 2004

Abstract The last two decades in the field of artificial intelligence have clearly shown that true intelligence always requires the interaction of an agent with a real physical and social environment. The concept of embodiment that has been introduced to designate the modern approach to designing intelligence has far-reaching implications. Rather than studying computation alone, we must consider the interplay between morphology, materials, brain (control), and the environment. A number of case studies are presented, and it is demonstrated how artificial evolution and morphogenesis can be used to systematically investigate this interplay. Taking these ideas into account requires entirely novel ways of thinking, and often leads to surprising results.

1 Introduction

In the traditional paradigm cognition, or generally intelligence, has been viewed as computation. The last two decades of research in the field have shown the limitations of this approach: true intelligence always requires the interaction with a real physical and social environment. An analysis of the failures of the traditional approach towards understanding and designing intelligent systems yields a fundamental neglect of the system–environment interaction. In contrast to a virtual or formal world (like chess, logic, or a virtual machine), the real world does not have precisely defined states; there is always only limited information available, there is only partial predictability, the environment has its own dynamics, and what an agent can do is not (completely) defined by the current situation. The interaction with the environment is always mediated by a physical body, with a particular morphology, i.e., body shape, and sensors and actuators distributed on the body. The concept of embodiment that has been introduced to designate the modern approach to designing intelligence has far-reaching implications. Rather than studying computation alone, we must consider the interplay between morphology, materials, brain (control), and the environment. These considerations go far beyond the trivial meaning of embodiment that "intelligence requires a body." They not only necessitate the interdisciplinary cooperation of computer science, neuroscience, engineering, and material science, but require entirely novel ways of thinking. It is interesting to note that agents do not "get" the information from the environment, but they have to actively acquire it through specific kinds of interactions, so called sensory–motor coordinations, as will be argued below.

As a first step toward a theory of intelligence based on the concepts of embodiment, a set of design principles for intelligent systems has been proposed which can be grouped into two categories, design procedure principles, and agent design principles. Examples of the former are "synthetic methodology," "time perspectives," "emergence," and "frame of reference," examples of the latter include "cheap design," "ecological balance," and "sensory–motor coordination." Because of their relevance to real-world interactions the focus will be on "cheap design," "ecological balance," and "sensory–motor coordination."

We start by summarizing the principles. We then pick out three principles for illustration. We then briefly outline how to systematically explore the design principles using artificial evolution and morphogenesis. To conclude, a number of research challenges and some speculations are presented. It should be noted that this is not a technical paper but a conceptual one.

R. Pfeifer (✉) · G. Gómez
Artificial Intelligence Laboratory, Department of Informatics at the University of Zurich, Andreasstrasse 15, CH-8050 Zurich, Switzerland
Tel. +41-1-635-4320/53; Fax +41-1-635-4507
e-mail: pfeifer@ifi.unizh.ch

This work was presented, in part, at the 9th International Symposium on Artificial Life and Robotics, Oita, Japan, January 28–30, 2004

Table 1. Overview of the design principles

Number	Name	Description
Design procedure principles		
P-Princ 1	Synthetic methodology	Understanding by building
P-Princ 2	Emergence	Systems designed for emergence are more adaptive
P-Princ 3	Diversity-compliance	Trade-off between exploiting the givens and generating diversity solved in interesting ways
P-Princ 4	Time perspectives	Three perspectives required: "here and now," ontogenetic, phylogenetic
P-Princ 5	Frame-of-reference	Three aspects must be distinguished: perspective, behavior vs. mechanisms, complexity
Agent design principles		
A-Princ 1	Three constituents	Task environment (ecological niche, tasks) and agent must always be taken into account
A-Princ 2	Complete agent	Embodied, autonomous, self-sufficient, situated agents are of interest
A-Princ 3	Parallel, loosely coupled processes	Parallel, asynchronous, partly autonomous processes; largely coupled through interaction with the environment
A-Princ 4	Sensory–motor coordination	Behavior sensory–motor coordinated with respect to target; self-generated sensory stimulation
A-Princ 5	Cheap design	Exploitation of niche and interaction; parsimony
A-Princ 6	Redundancy	Partial overlap of functionality based on different physical processes
A-Princ 7	Ecological balance	Balance in complexity of sensory, motor, and neural systems: task distribution between morphology, materials, and control
A-Princ 8	Value	Driving forces; developmental mechanisms; self-organization

2 Design principles: overview

There are different types of design principles. Some are concerned with the general "philosophy" of the approach. We call them "design procedure principles," as they do not directly pertain to the design of the agents, but more to the way of proceeding. Another set of principles deals more with the actual design of the agent. We use the qualifier "more" to express the fact that we are often not designing the agent directly, but rather the initial conditions and the learning and developmental processes, or the evolutionary mechanisms and the encoding in the genome, as we will elaborate later. A first version of the design principles was published at the 1996 conference on *Simulation of Adaptive Behavior*.[1] A more elaborate version was published in the book *Understanding Intelligence*.[2] The current overview will be very brief as a more extended version is in preparation.[3] Table 1 summarizes the design principles.

P-Princ 1: the synthetic methodology principle. The synthetic methodology "understanding by building" implies on the one hand constructing a model (computer simulation or robot) of some phenomenon of interest (e.g., how an insect walks, how a monkey is grasping a banana, or how we recognize a face in a crowd). On the other hand we want to abstract general principles (some examples are given below). The term "synthetic methodology" was adopted from Braitenberg's seminal book *Vehicles: Experiments in Synthetic Psychology*.[4]

P-Princ 2: the principle of emergence. If we are interested in designing adaptive systems, we should aim for emergence. Strictly speaking, behavior is always emergent, as it cannot be reduced to internal mechanisms only; it is always the result of a system–environment interaction. In this sense, emergence is not all or none, but a matter of degree: the further removed from the actual behavior the designer commitments are made, the more we call the resulting behavior emergent.

P-Princ 3: the diversity–compliance principle. Intelligent agents are characterized by the fact that they are on the one hand exploiting the specifics of the ecological niche, and on the other by behavioral diversity. In a conversation, we have to comply with the rules of grammar of the particular language, but then we can generate an infinite diversity of sentences. This principle, or trade-off, comes in many variations in cognitive science, i.e., the plasticity–stability tradeoff in learning theory,[5] assimilation–accommodation in perception,[6] and exploration–exploitation in evolutionary theory.[7]

P-Princ 4: the time perspectives principle. A comprehensive explanation of the behavior of any system must incorporate at least three perspectives: (a) state-oriented, i.e., the "here and now," (b) learning and development, i.e., the ontogenetic view, and (c) evolutionary, i.e., the phylogenetic perspective.

P-Princ 5: the frame-of-reference principle. There are three aspects to be distinguished when designing an agent: (a) the perspective, i.e., are we talking about the world from the agent's perspective, or the one of the observer, or the designer?; (b) behavior is not reducible to an internal mechanism; trying to do that would constitute a category error; and (c) any apparently complex behavior of an agent does not imply complexity in the underlying mechanism (for more detail, see Simon[8] and Seth[9]).

A-Princ 1: the three-constituents principle. This very often ignored principle states that when designing an agent, we have to consider three components: (a) the definition of the ecological niche (the environment), (b) the desired behaviors and tasks, and (c) the agent itself. The main point of this principle is that it would be a fundamental mistake to design the agent in isolation. This is particularly important because much can be gained by exploiting the physical and social environment.

A-Princ 2: the complete agent principle. The agents of interest are autonomous, self-sufficient, embodied, and situated. This view, although extremely powerful and obvious, is not very often considered explicitly.

A-Princ 3: the principle of parallel, loosely coupled processes. Intelligence is emergent from an agent–environment interaction based on a large number of parallel, loosely coupled processes that run asynchronously and are connected to the agent's sensory–motor apparatus.

A-Princ 4: the principle of sensory–motor coordination. All intelligent behavior (e.g., perception, categorization, memory) is to be conceived as sensory–motor coordination. This sensory–motor coordination, in addition to enabling the agent to interact efficiently with the environment, serves the purpose of structuring its sensory input. One of the powerful implications is that the problem of categorization is greatly simplified through interaction with the real world because the latter supports the generation of "good" patterns of sensory stimulation, where "good" means correlated and stationary (at least for a short period of time). One of the essential points here is that sensory stimulation is generated through interaction with the environment, which is a physical process, not a computational one.

A-Princ 5: the principle of cheap design. Designs must be parsimonious and exploit the physics and the constraints of the ecological niche. A trivial example is a robot with wheels which exploits the fact that the ground is mostly flat. Other examples are given below.

A-Princ 6: the redundancy principle. Agents should be designed such that there is an overlap of functionality of the different subsystems. Examples are sensory systems where, for example, the visual and the haptic systems both deliver spatial information, but they are based on different physical processes (electromagnetic waves vs. mechanical touch).

A-Princ 7: the principle of ecological balance. This principle consists of two parts, the first concerns the relation between the sensory system, the motor system, and neural control. Given a certain task environment, there has to be a match in the complexity of the sensory, motor and neural systems of the agent. The second is about the relation between morphology, materials, and control. Given a particular task environment, there is a certain balance or task distribution between morphology, materials, and control (e.g., Hara and Pfeifer[10] and Pfeifer[11]). Often, if the morphology and the materials are right, control will be much cheaper. Since we are dealing with embodied systems, there will be two dynamics, the physical one, or body dynamics, and the control, or neural dynamics, that need to be coupled (e.g., Ishiguro et al.,[12]).

A-Princ 8: the value principle. This principle is, in essence, about motivation. It is about why the agent does anything in the first place. Moreover, a value system tells the agent whether the result of an action was positive or negative (this is a very fundamental issue; there is no room for a comprehensive discussion here, but for a more detailed description see Edelman[13]).

Note that this set of principles by no means is complete. For example, a set of principles for designing evolutionary systems is currently under development.

3 Illustrations of embodiment

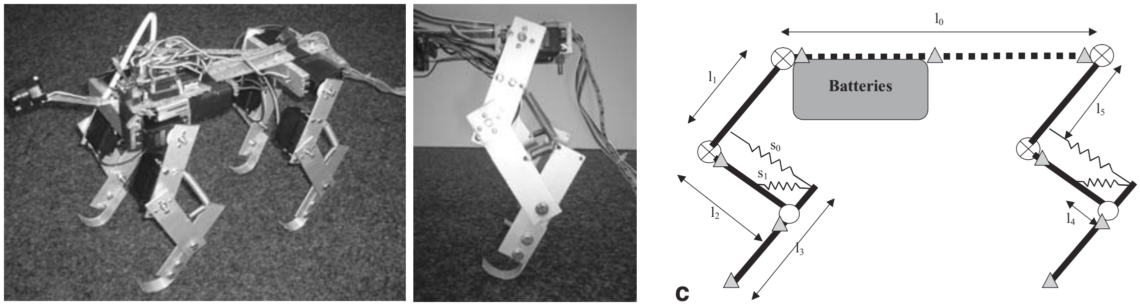
The following examples are to illustrate that embodiment has not only physical implications, but also important information theoretic ones (concerning neural control).

3.1 The passive dynamic walker, the quadruped "Puppy," and the dancing robot "Stumpy"

A passive dynamic walker is a robot (or, if you like, a mechanical device) capable of walking down an incline without any actuation and without control: it is "brainless," so to speak. In order to achieve this task, the passive dynamics of the robot, its body, and its limbs must be exploited. This kind of walking is very energy-efficient and there is an intrinsic naturalness to it. However, its "ecological niche" (i.e., the environment in which the robot is capable of operating) is extremely narrow: it only consists of inclines of certain angles. Energy-efficiency is achieved because in this approach the robot is, loosely speaking, operated near one of its Eigenfrequencies. To make this work, a lot of attention was devoted to morphology and materials. For example, the robot is equipped with wide feet of a particular shape to guide lateral motion, soft heels to reduce instability at heel strike, counter-swinging arms to negate yaw induced by leg swinging, and lateral-swinging arms to stabilize side-to-side lean.[14]

The quadruped "puppy" (Fig. 1), developed by Fumiya Iida of the AILab of the University of Zurich, represents another example of the exploitation of dynamics and of the interplay of morphology, materials, and control.[15,16]

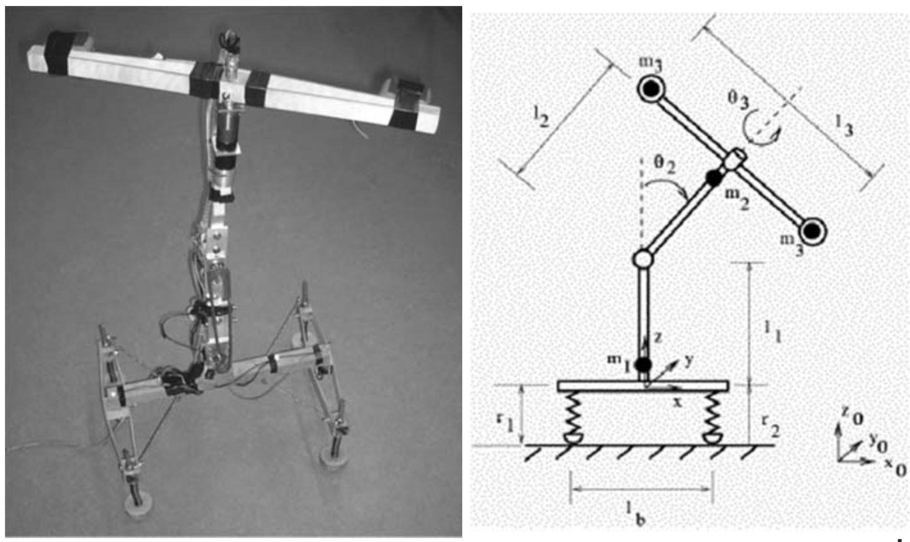
The legs perform a simple oscillation movement, but in the interaction with the environment, through the interplay of the spring system, the flexible spine (note that the battery is attached to as elastic spine which provides precisely the proper weight distribution), and gravity, a natural quadruped gait occurs, sometimes with all four legs off the ground. The system has self-stabilizing characteristics. It is interesting to note that the foot–ground contact must a,b



c

Fig. 1. The quadruped "puppy." **a** Picture of the entire "puppy." **b** The spring system in the hind legs. **c** Diagram showing the joints, servomotor actuated joints (*circles* with *crosses*), and flexible spine (*dotted line*).

Fig. 2. The dancing, walking, and hopping robot Stumpy. **a** Photograph of the robot. **b** Schematic drawing (for details, see text)



a b

exhibit little friction in order for this self-stabilization to work.

For "Stumpy,"[17,18] the goal was to generate a large behavioral diversity with as little control as possible. Stumpy's lower body is made of an inverted "T" mounted on wide springy feet (Fig. 2). The upper body is an upright "T" connected to the lower body by a rotary joint, the "waist" joint. The horizontal beam on the top is weighted on the ends to increase its moment of inertia. It is connected to the vertical beam by a second rotary joint, providing one rotational degree of freedom, in the plane normal to the vertical beam, the "shoulder" joint. Stumpy's vertical axis is made of aluminum, while both its horizontal axes and its feet are made of oak wood.

Stumpy can locomote in many interesting ways: it can move forward in a straight or curved line, it has different gait patterns, it can move sideways, and it can turn on the spot. Interestingly, this can all be achieved by actuating only two joints with one degree of freedom. In other words, control is extremely simple – the robot is virtually "brainless." The reason this works is because the dynamics, given by its morphology and its materials (elastic, spring-like materials, the surface properties of the feet), are exploited in clever ways.

These three case studies illustrate the principles of cheap design and ecological balance. Loosely speaking, we can say that the control tasks, i.e., the neural processing, are partly (or completely in the case of the passive dynamic walker) taken over by having the proper morphology and the right materials. Note that cheap design is not restricted to simple systems: it also applies to humans as highly complex biological creatures, as they also exploit the passive forward swing of the legs when walking.

3.2 Reaching and grasping: the principle of sensory–motor coordination as a key to higher levels of intelligence

Let us pursue this idea of exploiting the dynamics a little further and show how it can be taken into account to design actual robots. Most robot arms available today work with rigid materials and electric motors. Natural arms, by contrast, are built of muscles, tendons, ligaments, and bones, materials that are nonrigid to varying degrees. All these materials have their own intrinsic properties, like mass, stiffness, elasticity, viscosity, temporal characteristics, damping, and contraction ratio, to mention but a few. These properties are all exploited in interesting ways in natural systems. For example, there is a natural position for a human arm which is determined by its anatomy and by these properties. Reaching for and grasping an object like a cup with the right hand is normally done with the palm facing left, but could also be done (with considerable additional effort) the other way around. Assume now that the palm of your right hand is facing right and you let go. Your arm will immediately turn back into its natural position. This is not achieved by neural control, but by the properties of the muscle–tendon system (like a damped spring). Put differently, the morphology (the anatomy) and the materials provide physical constraints that make the control problem much easier – at least for the standard kind of movements.

There is an additional point of central interest. Assume that you are grasping an object. Through the act of grasping, a lot of rich sensory stimulation is generated at the finger tips, and because of the anatomy, the grasped object is (almost) automatically brought into the range of the visual system. Grasping, like pointing and reaching, are processes of sensory–motor coordination. Sensory–motor coordination is subtended by the anatomic (morphological) and material properties of the hand–arm–shoulder system, thus facilitating neural control. The sensory stimulation generated in this way implies correlations within and between sensory modalities, which is a prerequisite for developing higher levels of cognition. In this way, we are beginning to see how embodiment constitutes a precondition for intelligent behavior. The generation of structured sensory stimulation through physical interaction with the environment represents a key toward understanding developmental processes, as they are fundamental in humanoid robotics.

4 Artificial evolution and morphogenesis

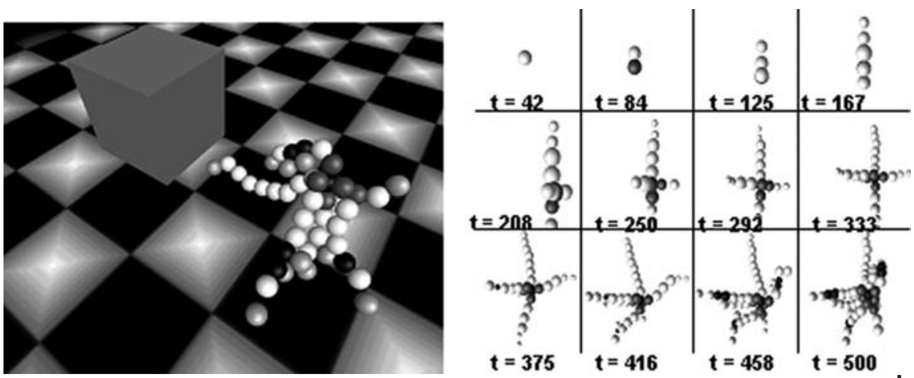
We have postulated and discussed a number of agent design principles. We also pointed out the principle of emergence. If we could demonstrate that the agent design principles would emerge from more fundamental evolutionary processes, this would corroborate the principles. As we are interested in embodied systems, we must define processes capable of co-evolving morphology, materials, and control. This can be achieved through artificial evolution with morphogenesis based on genetic regulatory networks. This way, we can study agent design systematically and observe the (potential) emergence of agent designs. In order to provide a feel for the methodology, we are including a paragraph with a short description of the mechanics of artificial genetic regulatory networks.

We provide a nontechnical introduction (for details, see, e.g., Bongard and Pfeifer[19] and Bongard[20,21]). It should be stressed that although this computational system is biologically inspired, it does not constitute a biological model. Rather, it is a system in its own right. Also, when we use biological terminology, e.g., when we say that “concentrations of transcription factors regulate gene expression,” this is meant metaphorically.

The basic idea is the following. A genetic algorithm is extended to include ontogenetic development by growing agents from genetic regulatory networks. In the example presented here, agents are tested for how far they can push a large block (which is why they are called “block pushers”). Figure 3a shows the physically realistic virtual environment. The fitness determination is a two-stage process: the agent is first grown, and is then evaluated in its virtual environment. Figure 3b illustrates how an agent grows from a single cell into a multicellular organism.

The algorithm starts with a string of randomly selected floating point numbers between 0 and 1. A scanning mechanism determines the location of the genes. Each gene consists of six floating point numbers which are the parameters that evolution can play with. These are explained in Fig. 4. There are transcription factors that only regulate the activity of other genes, and there are transcription factors for morphology and for neuronal growth. Whenever a gene is “expressed,” it will diffuse a transcription factor into the cell from a certain diffusion site. The activity of this genetic regulatory network leads to particular concentrations of the transcription factors to which the cell is sensitive: whenever a concentration threshold is exceeded, an action is taken.

For example, the cell may increase or decrease in size, if it gets too large it will split, the joint angles can be varied, neurons can be inserted, connections can be added or deleted, structures can be duplicated, etc. The growth process begins with a single unit into which “transcription factors” are injected (which determines the primary body axis). Then it is left to the dynamics of the genetic regulatory network. The resulting phenotype is subsequently tested in



a b

Fig. 3. Example of Bongard’s “block pusher.” **a** An evolved agent in its physically realistic virtual environment. **b** The growth phase starting from a single cell, showing various intermediate stages (last agent after 500 timesteps)

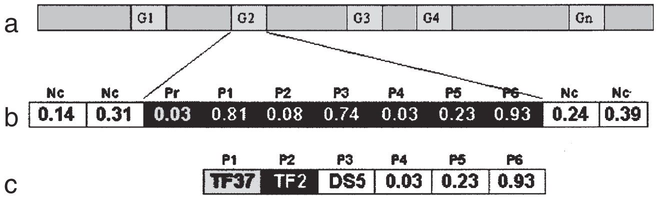


Fig. 4. The mechanisms underlying the genetic regulatory networks. **a** Genes on the genome. Which regions are considered to be genes is determined by an initial scanning mechanism (values below 0.1 are taken as starting positions). **b**, **c** An example of a particular gene. *Nc* means a "noncoding" region, *Pr* is a promoter site (the start of a gene), *P1* through *P6* are the parameters of the gene. *P1*, the transcription factor (*TF*) that regulates the expression of this gene [0,19]; *P2*, the *TF* the gene emits if expressed [0,42]; *P3*, the diffusion site, i.e., the location in the cell from which the *TF* is diffused; *P4*, the quantity of *TF* emitted by this gene, if expressed; *P5*, *P6*, lower and upper bounds of the concentrations within which the gene is expressed

the virtual environment. Over time, agents evolve that are good at pushing the block.

5 Conclusions: research challenges

Let us conclude by listing a few research challenges. (1) A theoretical understanding of (intelligent) behavior. In spite of half a century of research in artificial intelligence, we are still lacking a profound understanding of the mechanisms of intelligent behavior. With the set of design principles provided earlier, we hope to make a pertinent contribution, however small. At the moment, these principles are qualitative in nature, and a more quantitative formulation will be required in the future. (2) Achieving higher levels of intelligence through development. We touched only briefly upon sensory–motor coordination as a principle that is instrumental in achieving higher levels of intelligence. The field of "developmental robotics" capitalizes on this issue, and we can expect many exciting results from it. However, the field in its current state is lacking a firm theoretical foundation. (3) Fully automated design methods (artificial evolution and morphogenesis). One of the big challenges is the automation of design. Designing embodied systems presents an additional challenge, as we need to take into account the interplay between the environment (physical, but also social), morphology, materials, and control. (4) Moving into the real world (evolution, growth, etc.) To date, growth processes can only be achieved in simulation experiments – real-world growth processes are only in their very initial stages in research laboratories, and cannot yet be exploited for growing sophisticated creatures. This point represents an enormous challenge and will require many years of basic research.

Acknowledgments Rolf Pfeifer would like to thank Yasuo Kuniyoshi who initiated his visiting professorship at the 21st century COE at the School of Information Science and Technology of the University of Tokyo, and Masanori Sugisaka for inviting him to the AROB 9th Conference. For Gabriel Gómez funding has been provided by the EU-Project ROBOT-CUB: ROBotic Open-architecture Technology for Cognition, Understanding and Behavior (IST-004370).

References

1. Pfeifer R (1996) Building "fungus eaters": design principles of autonomous agents. In: Maes P, Mataric M, Meyer J-A, Pollack J, Wilson SW (eds) From animals to animats. 4. Proceedings of the 4th International. Conference on Simulation of Adaptive Behavior. MIT Press, Cambridge, p 3–12
2. Pfeifer R, Scheier C (1999) Understanding intelligence. MIT Press, Cambridge
3. Pfeifer R, Bongard J (in preparation). How the body shapes the way we think: the embodied revolution in artificial intelligence. MIT Press, Cambridge
4. Braitenberg V (1984) Vehicles: experiments in synthetic psychology. MIT Press, Cambridge
5. Grossberg S (1995) The attentive brain. Am Sci 83:438–449
6. Piaget J (1970) Piaget's theory. In: Mussen PH (ed) Carmichael's manual of child psychology. Wiley, New York, p 703–732
7. Eiben AE, Schippers CA (1998) On evolutionary exploration and exploitation. Fundamenta Informaticae 35:35–50
8. Simon HA (1969) The sciences of the artificial, 2nd edn. MIT Press, Cambridge
9. Seth AK (2002) Agent-based modeling and the environmental complexity thesis. In: Hallam J et al. (eds) Proceedings of the 7th International Conference on the Simulation of Adaptive Behaviour. MIT Press, Cambridge, p 13–24
10. Hara F, Pfeifer R (2000) On the relation among morphology, material and control in morpho-functional machines. In: Meyer, Berthoz, Floreano, Roitblat, Wilson (eds) From animals to animats. 6. Proceedings of the 6th International Conference on Simulation of Adaptive Behavior 2000, p 33–40
11. Pfeifer R (2003) Morpho-functional machines: basics and research issues. In: Hara F, Pfeifer R (eds) Morpho-functional machines: the new species. Springer, Tokyo
12. Ishiguro A, Ishimaru K, Hayakawa K et al. (2003) Toward a "well-balanced" design: A robotic case study. How should control and body dynamics be coupled? Proceedings of the 2nd International Symposium on Adaptive Motion of Animal and Machines
13. Edelman GE (1987) Neural Darwinism. The theory of neuronal group selection. Basic Books, New York
14. Collins SH, Wisse M, Ruina A (2001) A three-dimensional passive–dynamic walking robot with two legs and knees. Int J Robotics Res 20:607–615
15. Iida F, Pfeifer R (2004) "Cheap" rapid locomotion of a quadruped robot: self-stabilization of bounding gait. Intelligent Autonomous Systems vol 8. Groen F et al. (eds) IOS Press, p 642–649
16. Iida F, Gómez G, Pfeifer R (in press). Exploiting body dynamics for controlling a running quadruped robot
17. Paul C, Dravid R, Iida F (2002) Control of lateral bounding for a pendulum-driven hopping robot. Proceedings of the 5th International Conference on Climbing and Waling Robots (CLAWAR 2002), p 333–340
18. Iida F, Dravid R, Paul C (2002) Design and control of a pendulum-driven hopping robot. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS-2002, Lausanne, Switzerland, p 2141–2146
19. Bongard JC, Pfeifer R (2001) Repeated structure and dissociation of genotypic and phenotypic complexity in artificial ontogeny. In: Spector L et al. (eds) Proceedings of the 6th European Conference on Artificial Life, p 401–412
20. Bongard JC (2002) Evolving modular genetic regulatory networks. Proceedings of the IEEE 2002 Congress on Evolutionary Computation (CEC2002). MIT Press, Cambridge, p 305–311
21. Bongard JC (2003) Incremental approaches to the combined evolution of a robot's body and brain. Unpublished PhD thesis. Faculty of Mathematics and Science, University of Zurich