Active Perception: A Sensorimotor Account of Object Categorization

Stefano Nolfi[1]
[1]Institute of Cognitive Science and Technologies CNR,
Viale Marx, 15, 00137, Rome, Italy
nolfi@ip.rm.cnr.it

Davide Marocco[2-1]
[2] University of Calabria, "Centro Interdipartimentale della Comunicazione"
87036 Arcavacata di Rende, Cosenza, Italy
davidem@ip.rm.cnr.it

Abstract

We describe the results of a set of experiments in which we evolved the control system of artificial agents that are asked to categorize objects with different shapes on the basis of tactile information. Agents are provided with a 3-segments arm with 6 degrees of freedom and extremely coarse touch sensors. As we will see, despite such a limited sensory systems, evolved individuals are perfectly able to solve the problem. The analysis of the obtained results shows that evolved individuals always develop a well defined behavioral strategy that allows them to easily and robustly discriminate different objects despite the limitation of their sensory apparatus. Moreover, we discuss the general advantage of the evolutionary method for the synthesis of effective artificial agents.

1. Introduction

The behavior of embodied and situated organisms is an emergent result of the dynamical interaction between the nervous system, the body, and the external environment Ashby, 1952; Beer, 1995). This simple consideration has several important consequences that are far from being fully understood. One important aspect, for instance, is the fact that motor actions partially determine the sensory pattern that organisms receive from the environment. By coordinating sensory and motor processes organisms can select favorable sensory patterns and thus enhance their ability to achieve their adaptive goals.

Examples of processes falling within this category have been identified in natural organisms. Dill *et al.* (1993) demonstrated that since the fruit fly drosophila cannot always recognize a pattern appearing at different locations in the retina, the insect solves this problem of shift invariance by moving so to bring the pattern to the same retinal location where it has been presented during the storage process. Franceschini (1997) demonstrated that flies use motion to visually identify the depth of perceived obstacles. Moreover, there is evidence that environmental feedback obtained through motor actions plays a crucial role in normal development (Chiel and Beer, 1997; Thelen and Smith, 1994).

With few notable exceptions (eg, Braitenberg, 1984; Franceschini, 1997; Scheier and Pfeifer, 1995), the possibility of exploiting sensorimotor coordination in the design of artificial systems has largely been left unexplored. This can be explained by considering that, as we said above, behavior is the emergent result of the interactions between the individual and the environment. Given that in dynamical systems there is a complex and indirect relation between the rules that determine the interactions and the emergent result of those interactions, it is very difficult to identify how the interactions between the organism and the external environments contribute to the resulting behavior. As a consequence, designing systems that exploit sensorimotor coordination is rather difficult (for an attempt to identify new design principles that might help to achieve this goal, see Pfeifer and Scheier [1999]).

From this point of view evolutionary experiments (Nolfi and Floreano, 2000) where artificial organisms autonomously develop their skills in close interaction with the environment represent an ideal framework for studying sensorimotor coordination (Nolfi and Floreano, 2002). Indeed, in most of the experiments conducted with artificial evolution one can observe the emergence of behavior exploiting active perception. The analysis of evolved robots 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. In addition, such analysis may allow the identification of new explanatory hypotheses that may produce new models of natural behavior that, later on, might be tested experimentally on the real organisms.

In the next section we describe the results of a set of experiments in which we evolved the control system of artificial agents that are asked to categorize objects with different shapes on the basis of tactile information. Agents are provided with a 3-segment arm with 6 degrees of freedom and extremely coarse touch sensors. As we will see, despite such a limited sensory systems, evolved individuals are perfectly able to solve the problem. The analysis of the obtained results shows that evolved individuals always develop a well defined behavioral strategy that allows them to easily and robustly discriminate different objects despite the limitation of their sensory apparatus. Finally, in the conclusions we discuss the general advantage of the evolutionary method for the synthesis of effective artificial agents.

2. Experimental Results

We evolved the control system of agents that are asked to categorize objects with different shapes on the basis of tactile information. Agents are provided with a 3-segments arm with 6 degrees of freedom (DOF) and extremely coarse touch sensors (see Figure 1).

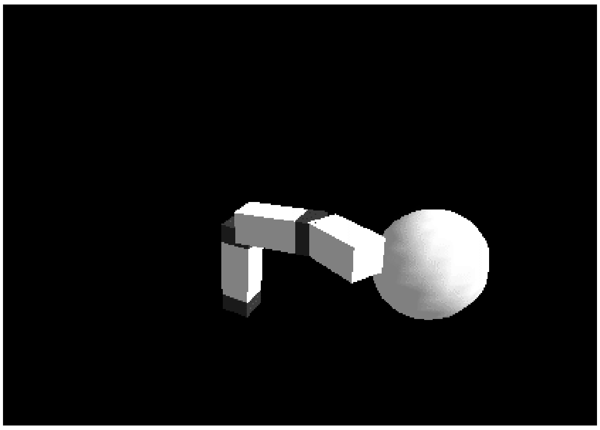

Figure 1. The arm and a spherical object.

To reduce the time necessary to test individual behaviors and model the real physical dynamics as accurately as possible we used the rigid body dynamics simulation SDK of Vortex^{TM} (see http://www.cm-labs.com/products/vortex/). This software allowed to build our robotic arm by means of several segments connected by joints and to run simulations faster than real time.

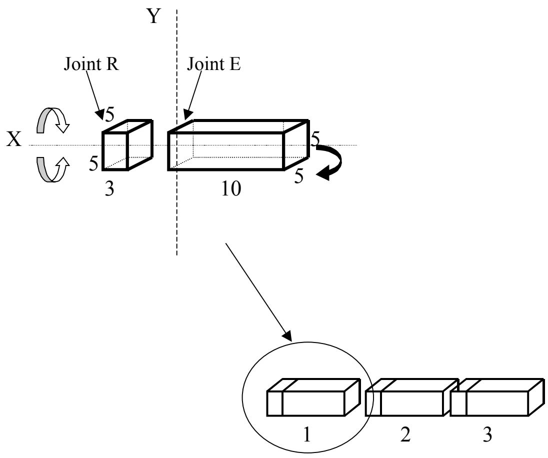

Figure 2. A schematic representation of the arm.

Given the specific characteristics of this tool, the implementation of the arm consists of a basic structure comprising two segments and two joints replicated three times (see Figure 2). The basic structure consists of a shorter segment of size {x=5, y=3, z=5} and a longer segment of size {x=5, y=10, z=5}. This two segments are connected by means of a joint (i.e. the *Joint E* in the Figure) that allows only one DOF on axis *Y* , while the shorter segment is connected at the floor, or at the longer segment, by means of a joint (i.e. the *Joint R*) that provides one DOF on axis *X*. In practice, the *Joint E* allows to elevate and lower the connected segments and the *Joint R* allows to rotate them in both direction. Notice that *Joint E* is free to moves only in a range between 0 and $\pi/2$, just like an human arm that can bend the elbow solely in a direction. The range of *Joint R* is [$-\pi/2$, $+\pi/2$] Gravity is {0, -1,0, 0}. Each actuator is provided with a corresponding motor that can apply a maximum force of 50. Therefore, to reach every position in the environment the control system has to appropriately control several joints and to deal with the constraints due to gravity. Friction was set to 2.0.

The sensory system consists of a simple contact sensor placed on each longer segment that detects when this segment collides with an other object and proprioceptive sensors that provide the current position of each joint.

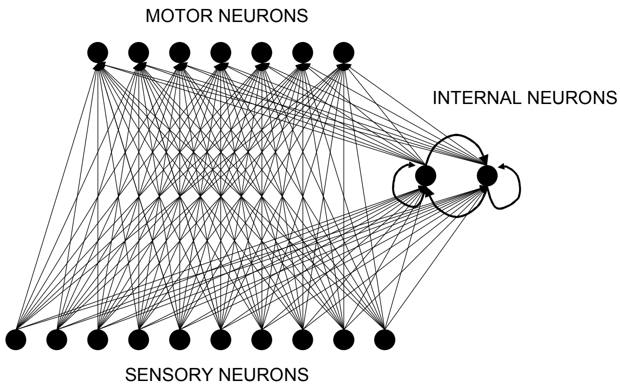

Figure 3. The architecture of the neural controller.

Each individual of the population was tested for 36 phases, each phase consisting of 150 timesteps. At the beginning of each phase the arm is fully extended and a spherical or a cubic object is placed in a random selected position in front of the arm (the position of the object is randomly selected between the following intervals: 20.0 >= X <= 30.0; 7.5 >= Y <= 17.5; -10.0 >= Z <= 10.0). The object is a sphere (15 units in diameter) during even phases and a cube (15 units in side) during odd phases so that each individual has to discriminate the same number of spherical and cubic objects during its "lifetime".

The controller of each individual consists of a neural networks with 10 sensory neurons directly connected to 7 motor neurons and 2 internal neurons receiving connections from the sensory neurons and from themselves and projecting connections to the motor neurons (see Figure 3). The first 9 sensory neurons encode the angular position (normalized between 0.0 and 1.0) of the 6 DOF of the joints and the state of the three contact sensors located in the three corresponding segments of the arm. The last sensory neuron is a copy of the last motor neuron that encode the current classification produced by the individual (see below). The first 6 motor neurons control the actuators of the 6 corresponding joints. The output of the neurons is normalized between [0, +π/2] and [-π/2, +π/2] in the case of elevation or rotational joints respectively and is used to encode the desired position of the corresponding joint. The motor is activated so as to apply a force (up to 50 units) proportional to the difference between the current and the desired position of the joint. The seventh motor neuron encodes the classification of the object produced by the individual (value below or above 0.5 are interpreted as classifications corresponding to a cubic or spherical object respectively).

The activation state of sensory and internal neurons was updated accordingly to the following equations (motor neurons were updated according to the logistic function):

$$A_j = t_j + \sum w_{ij} O_i$$
$$O_j = \tau_j O_j^{(t-1)} + (1-\tau_j)(1+e^{-A_j})^{-1} \qquad \textbf{(1)}$$
$$0 \le \tau_j \le 1$$

With *Aj* being the activity of the *j*th neuron (or the state of the corresponding sensor in the case of sensory neurons), *tj* the bias of the *j*th neuron, *Wij* the weight from the *i*th to the *j*th neuron, *Oi* the output of the *i*th neuron. *Oj* is the output of the *j*th neuron, *τj* the time constant of the *j*th neuron. It should be noted that similar, although slightly worse performance, was obtained by using the standard logistic function for all neurons (result not shown).

The genotype of evolving individuals consists of 139 parameters that include 108 weights, 19 biases, and 12 time constants. Each parameter is encoded with 8 bits. Weights and biases are normalized between –10.0 and 10.0, time constants are normalized between 0.0 and 1.0. The fitness of individuals is computed by summing the number of phases in which the individuals correctly classify the current object. The classification is correct if at the end of the phase (i.e. after 150 cycles) the activation of the last motor units is below 0.5 and the object is a cube or is above 0.5 and the object is a sphere. By running 10 replications of the experiment and by evolving individuals for 50 generations we observed that in many of the replications evolved individuals display a good ability to classify the two objects and, in some cases, they produce close to optimal performance. Figure 4 shows the percentage of correct classifications measured through 100 trials for the best individual of each generation. As can be seen, in the case of the best replication (thin line), evolved individuals reach up to 98% of correct classifications.

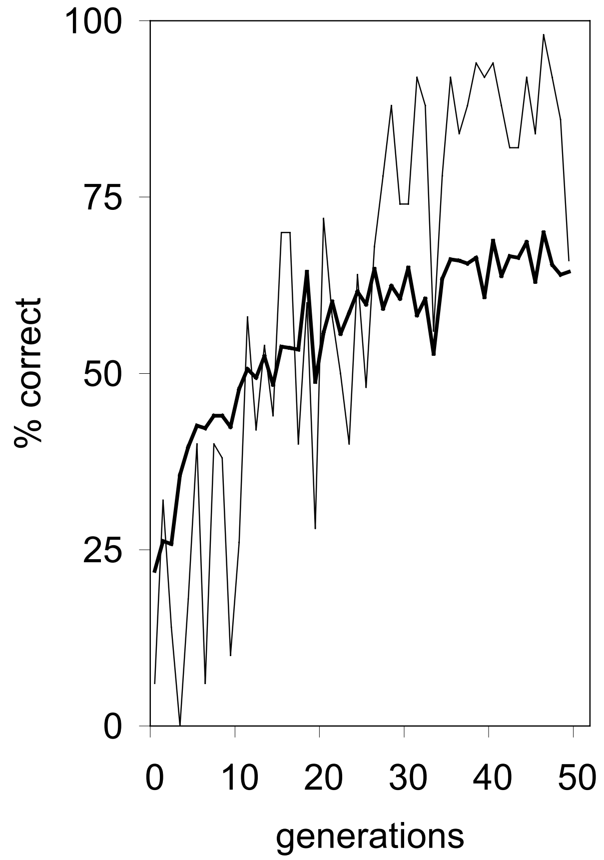

Figure 4. Percentage of correct classifications against generation for the best individual of each generation. The tick line represents the average performance of 10 replications. The thin line represents the performance of the best replication. Each individual has been tested for 100 phases.

By analyzing the obtained behaviors one can clearly see that in all experiments evolved individuals select a well defined behavior that assures that perceived sensory states corresponding to different objects can be easily discriminated and allows robust and effective categorizations. Figure 5 shows how a typical evolved individual behaves with a spherical and a cubic object (left and right side of the Figure respectively). As can be seen, first the arm bends on the left side and move to the right in order to start to feel the object with the touch sensor of the third segment. Then the arm moves so as to follow the curvilinear surface of the sphere or to keep touching one of the angles of the cubic object.

The fact that such behavior significantly simplifies the discrimination of the two objects can be explained by considering that the arm ends in very different conditions in the case of a sphere or of a cubic object. In particular, after a certain amount of time in which the arm is negotiating the object, it ends almost fully extended in the case of a spherical object and almost fully bent in the case of a cubic object. This implies that, given such a behavior, the state of the proprioceptive sensors after a certain amount of time can be used as a direct and straightforward indication of the category of the object. The fact that such behavior allows evolved individuals to produce robust and effective classifications can be explained by considering that the final classification is not the result of a single decision but is the end result of an interaction between the agent and the object that last several timesteps during which the agent keeps following the surface of the object so to ascertain whether it is curvilinear or not. Indeed, evolved individuals that display shorter negotiation periods with spherical objects also produce worse classification performance (result not shown).

The analysis of the activation state of the neurons during the behavior displayed in Figure 5 (see graphs H1 and H2 in Figure 6 and 7 that show the activation of the internal neurons when the arm has to discriminate the spherical or the cubic object respectively) shows that internal units are activated during the first phase (when the arm is looking for the object) and not activated during the second phase (when the arm starts to negotiate the object) for both spherical and cubic objects. Also notice how the activation state of unit C, that encodes the classification produced by the agent, starts low and then increases when the arm negotiates the sphere (Figure 6) while starts and remains low when the arm negotiates the cube (Figure 7). The fact that, at the end of the phase, the internal units tend to have the same activation states in the two cases shows that the classification is not accomplished on the basis of internal information extracted during the interaction between the arm and the object but rather on the basis of the final position of arm itself that, as claimed above, directly provides a clear indication of the category of the object with which the agent has previously interacted.

3. Discussion

Passive approaches to perception (e.g. Shapiro, 1987) assume that perception consists of the construction of a detailed representation of the external world. From this point of view the main challenge is that of transforming egocentric, incomplete, and noisy sensory information into an allocentric, complete, and precise representation of the external environment. To achieve this goal a large number of hard problems (in the case of vision, for example, infer 3D surfaces from 2D images or handle occlusions) have to be solved. Perception thus typically involve an intensive static analysis of passively sampled data. Within this view, motor behavior (i.e. the interaction with the external world) is not viewed as an opportunity but rather as a problem to be controlled --- the result of the perceptual process should be as independent as possible from the behavior displayed by the agent during the collection of sensory data.

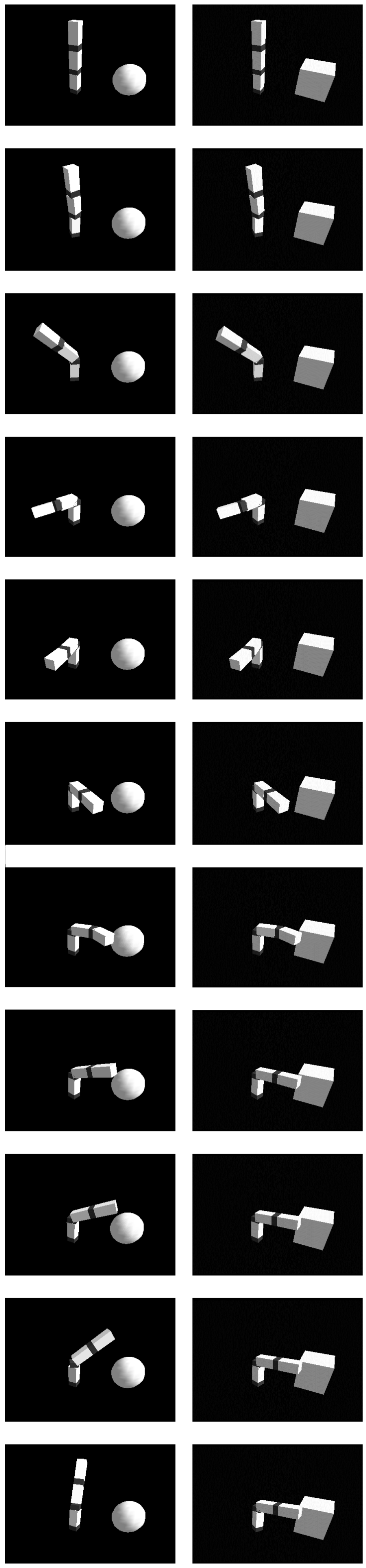

Figure 5. Behavior of a typical evolved individual during an phase (150 cycles) in which the object consists of a sphere (left pictures) and of a cube (right pictures). For reason of space, the pictures show the position of the arm each 15 cycles.

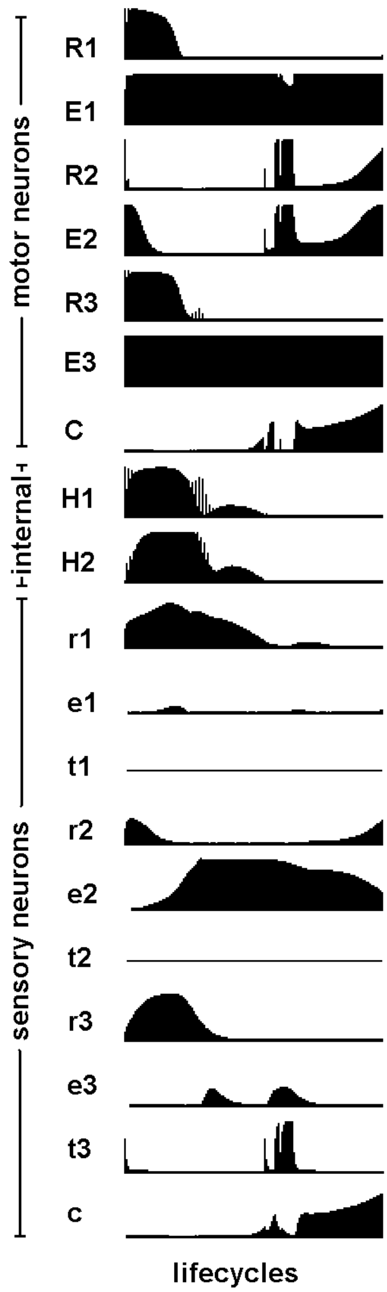

Figure 6. Activation state of the neurons during the behavior shown in the left side of Figure 4 through out 150 cycles. The height with respect to the baseline represents the activation state of the unit. *R1-R3* and *E1-E3* represent the activation state of the motor units that control the rotation and the elevation respectively of the three corresponding joints. *C* is the classification unit (value below and above 0.5 corresponds to spherical and cubic objects respectively). *H1* and *H2* represent the activation state of the two internal neurons. *r1-r3* and *e1-e3* represent the activation state of the sensory neurons that encode the current rotation and elevation of the three corresponding segments. *t1-t3* represent the activation state of the touch sensors located on the three corresponding segments. *c* is the copy of the *C* classification unit.

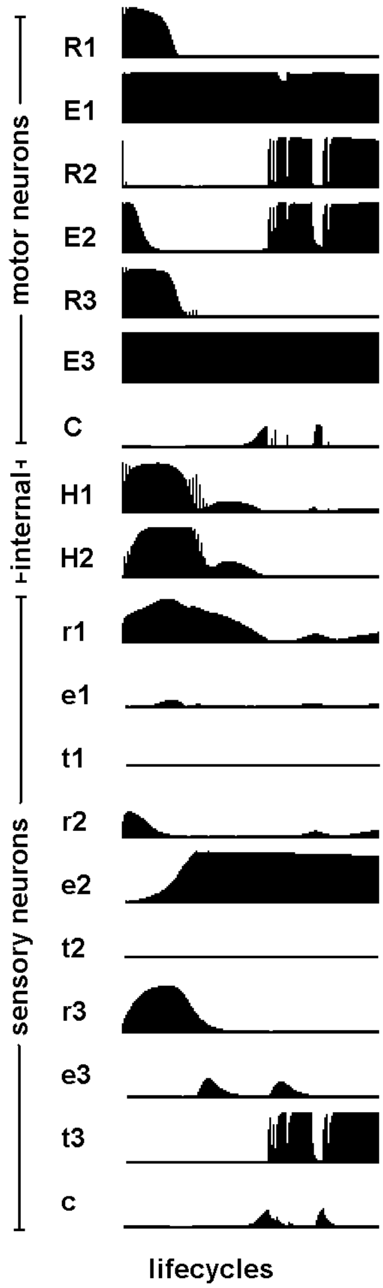

Figure 7. Activation state of the neurons during the behavior shown in the left side of Figure 4 through out 150 cycles. See legend of Figure 6.

Active approaches to perception (Bajcsy, 1988; Ballard, 1991), on the contrary, assume that the outside world serves as its own, external, representation and perception consists in mastering the regularities arising from sensorimotor interactions. From this point of view perception is a way of acting or, in other words, an exploratory activity of the environment. As pointed out by O'Reagan (2001, pp. 3) perception consists in identifying "the structure of the rules governing the sensory changes produced by various motor actions”. Within this view perception and action cannot be separated and the produced behavior plays a crucial role in the outcome of the perception process.

From an engineering point of view active perception has several advantages. In many cases, solutions that exploit active perception, in fact, are much simpler than solutions that rely on passive approaches to perception (Bajcsy, 1988; Ballard, 1991; O'Reagan, 2001). In addition, active approaches, by not relying on a detailed internal representation of the external environment, are less affected by the problem of how to update such an internal representation when the environmental conditions change. On the other hand, active approaches require the designer to identify the appropriate behavior that in turn allows the agent to identify sensorimotor regularities that provide useful information. This task --- namely the identification of the appropriate way of interacting with the environment --- may be extremely difficult from the point of view of the designer given that, as we claimed in the introduction, behavior is the emergent result of the interaction between the agent and the environment. Therefore, from the point of view of the human designer that has to manually program the agent, the advantages of active perception might be counterbalanced by the difficulties of programming effective behaviors.

As we showed in this paper, evolutionary techniques in which individual agents are selected on the basis of the overall behavior emerging from the interaction between their control system and the environment (Nolfi and Floreano, 2000) represent an effective way to develop systems that are able to exploit active perception and, at the same time, to release the designer from the burden of identifying and programming the appropriate exploratory behaviors. The fact that similar results have been obtained by evolving wheeled robots, provided with different sensory systems ranging from infrared sensors to visual cameras, asked to categorize different type of objects (Scheier, C. *et al.*, 1998; Nolfi and Marocco, 2000; Nolfi 2002) demonstrates that the evolutionary method has a general validity and can be successfully applied to tackle different problems.

References

Ashby, W.R. (1952) *Design for a Brain*. London: Chapman and Hall

Bajcsy, R. (1988) Active perception *Proceedings of the IEEE* (76) 8, 996-1005

Ballard D.H. (1991) Animate vision. *Artificial Intelligence* 48:57-86.

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

Braitenberg, V. (1984) *Vehicles*. Cambridge, MA: MIT Press

Chiel, H.J. and 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

Dill, M. *et al*. (1993) Visual pattern recognition in Drosophila involves retinotopic matching *Nature* 365, 751-753

Franceschini, N. (1997) Combined optical, neuroanatomical, electrophysiological and behavioral studies on signal processing in the fly compound eye. In C. Taddei-Ferretti (Ed.), *Biocybernetics of Vision: Integrative Mechanisms and Cognitive Processes*. London: World Scientific

Nolfi, S. (2002) Power and Limits of Reactive Agents. *Neurocomputing* 42:119-145

Nolfi, S. and Floreano, D. (2000) *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press

Nolfi S. and Floreano D. (2002) Synthesis of autonomous robots through artificial evolution, *Trends in Cognitive Sciences* 1:31-37.

Nolfi, S. and Marocco D. (2000) Evolving visually-guided robots able to discriminate between different landmarks. In: J-A Meyer, A. Berthoz, D. Floreano, H.L. Roitblat, and S.W. Wilson (eds.) *From Animals to Animats 6. Proceedings of the VI International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press. pp. 413-419

O'Regan, J.K. and Noë A. (2001) A sensorimotor account of vision and visual consciousness. *Behavior and Brain Sciences*, 5: ?-?.

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

Scheier, C. and Pfeifer, R. (1995) Classification as sensorimotor coordination: A case study on autonomous agents. In F. Moran, A. Moreno, J.J. Merelo and P. Chacon (Eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life*. Berlin: Springer Verlag

Scheier, C. *et al*. (1998) Embedded neural networks: exploiting constraints. *Neural Networks* 11, 1551-1596

Shapiro S. (1987) *Encyclopedia of Artificial Intelligence*. New York: Wiley Press.

Thelen, E. and Smith, L.B. (1994) *A Dynamics Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press