Basic Principles of Adaptive Learning through Variation and Selection

Mikhail S. Burtsev[1,2]

[1] Keldysh Institute of Applied Mathematics of RAS, 4 Miusskaya sq., Moscow, 125047, Russia
[2] P.K. Anokhin Institute of Normal Physiology of RAMS, 11/4 Mokhovaya, Moscow, 125009, Russia
mbur@ya.ru

Abstract

The evolutionary theory relies on the principles of variation and selection to explain adaptation. It is reasonable to fit these powerful principles to the learning theory. A number of selectionist approaches were proposed but found modest recognition so far. This theoretical paper attempts to review an application of basic ideas of the evolutionary adaptation to the lifetime learning. The analysis demonstrates that an adaptive value can be translated from the level of evolution to the level of individual through the innate repertoire of behaviors. This primary repertoire forms initial attractor for the behavioral dynamics. Learning starts when an environment offsets an organism from the existing attractor trajectory. Blind variations of behavior are generated until the return to the target attractor. These variations are retained to make up new branches of basin of attraction. It is important that the existed behavioral trajectory should not be altered as the learning unfolds because it keeps knowledge about adaptations survived selection through the evolutionary and learning history.

Introduction

Animals learn and this learning is usually beneficial or at least neutral for their evolutionary success. Generally, an adaptive learning is considered to be driven by some value system. The value system categorizes states of an environment in terms of their adaptive value. This categorization results in the feedback used in modification of behavior during learning. Adoption of the value system to explain the adaptive value of learning is not an exclusive solution. This paper addresses an application of the evolutionary principles of variation and selection to the explanation of adaptive outcome of learning. In spite of a number of selectionist theories of learning being proposed (Skinner, 1981; Edelman, 1987; Changeux and Dehaene, 1989) none of them gained widespread recognition. Here I try to analyze and clarify some basic ideas behind the selectionist approach. In particular, I focus on the issues concerning the initiation and finalization of learning, the selection criteria for the behavioral modifications, and memory retention.

An explanation of learning adaptive value is the ultimate problem of learning theory. Learning is adaptive when it leads to the modification of behavior that is evolutionary beneficial. But natural selection operates on the scale of generations and learning unfolds on the interval of minutes or even seconds. The evolutionary values should be transferred to the level of learning. This transfer is maintained by Darwinian evolution of developmental processes. And when an organism starts learning it already has criteria of adaptivness created during ontogeny.

Having representations of evolutionary values on the organismic level is a half of the story. The other half is generation of new adaptive behaviors. During learning an individual should produce behaviors taking into account evolutionary values. A mainstream of modern theories of animal learning in the fields of neuroscience (Schultz and Dickinson, 2000; Suri et al., 2001; Dayan and Balleine, 2002; Berridge and Robinson, 2003) and adaptive behavior (Maes, 1994; Sutton and Barto, 1998; Dorigo and Colombetti, 1998; Adaptive Behavior, 2002) employs “feedback” logic for aligning behavior with values. In this logic, discrepancy in the expected value guides learning. The value of an error signal is used to produce modifications of behavior.

An alternative approach to the generation of adaptation is presented by the explanatory scheme of the evolutionary theory. Evolution requires two processes, the first is generation of variation and the second is selection. The logic of the evolutionary explanatory scheme is *opposite* to the “feedback” logic mentioned above. In the “feedback” logic the adaptivness is evaluated first in terms of “reward” expectation mismatch and then the obtained error signal is used to change behavior. On the other hand, in the evolutionary scheme the generation of possible solutions goes first and then evaluation takes place in the form of selection. This reversal of stages leads to the next important distinction. The process of variation generation precedes the evaluation so it is *independent* of selection criteria, but in the “feedback” approach modifications *depend* on evaluations. The differences are outlined in the table 1.

“feedback” logic	**evolutionary logic**
evaluation then modification	generation then selection
modification depends on evaluation	generation is independent of selection

Table 1: The differences between “feedback” and “evolutionary” logic of adaptation.

It is obvious that the evolutionary logic of adaptation was successfully applied in a numerous studies to the synthesis of adaptive agents (Beer, 1996; Harvey et al., 1997; Pfeifer and Scheier, 1999; Nolfi and Floreano, 2000; Beer, 2000; Harvey et al., 2005) but its application to the problem of individual learning is still on the way.

There are a number of attempts to use principles of variation and selection in the theories of learning. In 1960 W. Ross Ashby published his influential book "Design for a brain" (Ashby, 1960) where one can find a proposal of the cybernetic theory of learning that utilizes trials and errors. Ashby had introduced so called essential variables (variables indicating viability of an organism) and treated them as a source of control for the blind variation. "Design for a brain" is focused mainly on the issue of behavior's stability. Here the important achievement was that adaptation in one behavioral subsystem should not disturb other subsystems of an animal and, hence, the subsystems for different behaviors should be loosely connected. Recently the similar idea of the structural adaptation was studied by Toussaint (Toussaint, 2004). Unfortunately, Ashby said nothing on retention of previous experience of an agent, and in his scheme of learning only the last successful adaptation is conserved. Therefore, each new learning episode should start from scratch and this would lead to the repetition of previous errors and make learning less effective. Also, Ashby allowed only a fixed set of essential variables to control blind variations.

On the conceptual level the evolutionary approach to learning lies at the intersection of evolutionary epistemology (Campbell, 1974; Popper, 1984) and constructivism (Glasersfeld, 1995; Foundations of Science, 2001). The famous Popperian formula describing the growth of knowledge through variation and selection is:

$$P_1 \to TT \to EE \to P_2, \qquad (1)$$

here P_1 stands for the initial problem, TT are tentative theories or solutions proposed to solve it, EE is a process of error elimination, and P_2 is a new problem. For the sequential scheme of generation of solutions formula is extended and takes form:

$$P_1 \to TT_1 \to EE_1 \to TT_2 \to EE_2 ... \to TT_n \to success, \qquad (2)$$

here n is a number of attempts which an agent has performed until solution was obtained.

An important contribution of the evolutionary epistemology is a concept of "vicarious selector" (Campbell, 1974). Vicarious selectors serve as internal representations of external factors of natural selection thus allowing transfer of evolutionary values to the level of learning. Vicarious selection "substitutes" natural selection during lifetime learning. The hierarchy of vicarious selectors accumulates all previous (even unsuccessful) experience and, as a consequence, generation of new behaviors takes the form of progressive growth on the top of existing competence. This process can be interpreted as a construction of individual knowledge by an active agent.

B.F. Skinner advocated his theory of "selection by consequences" (Skinner, 1981). Skinner considered selection by consequences as an explanatory scheme that is common for the three different levels, namely, the Darwinian evolution, learning and social evolution. Unfortunately, radical rejection of any attempts to consider the processes underling selection for all these cases made his approach fruitless.

There are two selectionist theories in the field of neuroscience: the theory developed by Changeux and Dehaene (Changeux and Dehaene, 1989), and the theory of neuronal groups selection (TNGS) proposed by Edelman (Edelman, 1987, 1993). Both theories declare thoroughly application of the "neural Darwinism" to the processes on all levels of brain organization from the synapse to the consciousness. The suggested sources of variation are generation of excessive synaptic connections during development and variable activation of neural assemblies. In their basic form both theories attributes selection to the input matching:

> "At a given stage of the evolution of the organism, some of these spontaneously generated pre-representations may not match any defined feature of the environment (or any item from long-term memory stores) and may thus be transiently meaningless. But, some of them will ultimately be selected in novel situations, thus becoming "meaning full". The achievement of such adequacy (fitness) with the environment (or with a given cognitive state) would then become the basic criterion for selection." (Changeux and Dehaene, 1989, p. 87)

The similar idea for the TNGS is presented in (Izhikevich et al., 2004). But input matching only is not enough for the creation of adaptive behavior because the action selection process should also be specified. The solution to this problem was put forward in (Dehaene and Changeux, 2000):

> "The models that we have introduced thus implement a generalized variation/selection scheme which was initially explored under the name of 'reinforcement learning' by computer scientists (e.g. Sutton and Barto, 1998) and has also been called 'neural Darwinism' by neurobiologists (Edelman, 1987, 1993; Changeux and Dehaene, 1989)."

This solution creates confusion because "generalized variation/selection scheme" refers to the evolutionary explanatory scheme but used by authors as a shortcut to the principles primary to reinforcement learning. The mechanisms of adaptation in the reinforcement learning fall in the domain of "feedback" logic which is opposite to the evolutionary one (see Table 1).

An integration of theoretical proposals devoted to application of the evolutionary principles to learning gives only some features of the required picture. Below an attempt to analyze the process of individual learning in the framework of evolutionary logic is presented and some consequences are discussed in relation to the current theoretical landscape.

Formalization

Below an organism or a robot is considered as an abstract adaptive agent. The agent's "brain" can be represented as an automaton **A** (similar approach was used by Peschl (Peschl, 1997) in his investigation of representations in neural systems). Description of the automaton **A** should include definitions of a set of all the components ***C*** of **A** and a transformation of these components states B that generates the "brain" dynamics.

The set of components ***C*** of the brain-automaton **A** consists of the following subsets (fig. 1):

- $\boldsymbol{C_E}$ is a subset of the components which states are determined by the external environment.
- $\boldsymbol{C_I}$ is a subset of components which states are determined by the internal state of the agent (its body).
- $\boldsymbol{C_S}$ is a subset of components which states are not determined directly by the external environment or the internal state of the agent but determined by any other components of the set $\boldsymbol{C}$.
- $\boldsymbol{C_A}$ is a subset of components which states determine the actions executed by the agent.

Hence a "brain" of an agent is specified by the automaton $\mathbf{A}\{\boldsymbol{C_E},\boldsymbol{C_I},\boldsymbol{C_S},\boldsymbol{C_A},B\}$.

$$\mathbf{A}\{\boldsymbol{C_E},\boldsymbol{C_I},\boldsymbol{C_S},\boldsymbol{C_A},B\}$$

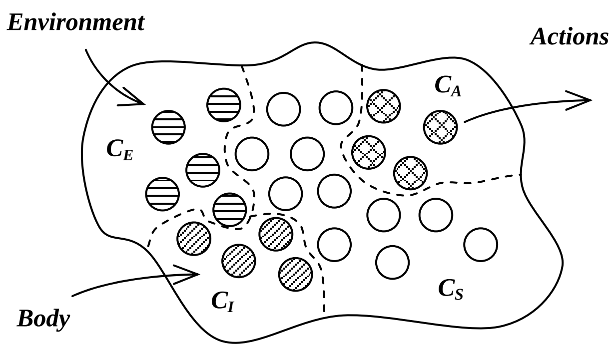

Figure 1: The automaton representation of the agent's "brain" (for details see text).

With the use of notation introduced above a behavior of the agent in discreet time $t\{t_0,t_1,t_2,\ldots,t_n\}$ is represented as modification of it's states by transformation B and can be written like this:

$$\ldots\boldsymbol{C}^{t_n} \xrightarrow{B} \boldsymbol{C}^{t_{n+1}} \xrightarrow{B} \boldsymbol{C}^{t_{n+2}} \xrightarrow{B} \ldots \tag{3}$$

or equivalently

$$\boldsymbol{C}^{t_{n+1}} = B(\boldsymbol{C}^{t_n}) = B(\boldsymbol{C}^{t_n}_{\boldsymbol{E}}, \boldsymbol{C}^{t_n}_{\boldsymbol{I}}, \boldsymbol{C}^{t_n}_{\boldsymbol{S}}, \boldsymbol{C}^{t_n}_{\boldsymbol{A}}), \tag{4}$$

here $\boldsymbol{C}^{t_n}$ is a vector of states of the automaton's components at the time t_n, or in other words the state of $\mathbf{A}$ at t_n.

The behavior of the agent is *constant* if the transformation B doesn't change with time. It should be mentioned that according to the equation (4) constancy of behavior is not necessary leads to the same actions in the same environmental and bodily conditions because the next state of the automaton also depends on the states of its internal components $\boldsymbol{C_S}$. It is reasonable to define learning as a transformation L of the brain dynamics B of the agent. Thus L is a transformation defined on the set of possible B's. To introduce learning into the dynamics the both B and L are applied:

$$\ldots\boldsymbol{C}^{t_n} \xrightarrow{L(B^{t_n})} \boldsymbol{C}^{t_{n+1}} \xrightarrow{L(B^{t_{n+1}})} \ldots, \tag{5}$$

or

$$\begin{cases} \boldsymbol{C}^{t_{n+1}} = B^{t_n}(\boldsymbol{C}^{t_n}) = B^{t_n}(\boldsymbol{C}^{t_n}_{\boldsymbol{E}}, \boldsymbol{C}^{t_n}_{\boldsymbol{I}}, \boldsymbol{C}^{t_n}_{\boldsymbol{S}}, \boldsymbol{C}^{t_n}_{\boldsymbol{A}}) \\ B^{t_{n+1}} = L(B^{t_n}) \end{cases}. \tag{6}$$

Addition of the learning transformation L to the automaton $\mathbf{A}$ results in an *automaton with learning* $\mathbf{A}\{\boldsymbol{C_E},\boldsymbol{C_I},\boldsymbol{C_S},\boldsymbol{C_A},B,L\}$.

"Feedback" logic of learning

The common assumption about logic of learning is that change in the function which generates a behavior (i.e. B) *is determined* by the states of the agent's "brain". These states can be external reinforcing stimuli which the agent percepts through the activation of some sensory inputs (i.e. $\boldsymbol{C_E}$), or the signals carrying information about the state of the body (i.e. $\boldsymbol{C_I}$), or the activations of pre- and post-synaptic neurons (i.e. $\boldsymbol{C_S}$) in the activity-dependent plasticity. In the framework of automata approach accepted in the paper this means that the transformation L is a function of values of states $\boldsymbol{C}$:

$$L = f(\boldsymbol{C_E}, \boldsymbol{C_I}, \boldsymbol{C_S}, \boldsymbol{C_A}). \tag{7}$$

In other words, the transformation from B^{t_n} to $B^{t_{n+1}}$ is determined by the state $\boldsymbol{C}^{t_n}$ of the automaton $\mathbf{A}$.

Evolutionary logic of learning

Following the evolutionary logic the adaptation is produced by variation and selection where generation of variation is independent of selection. One can start with the most radical assumption that a change in the function which generates behavior (i.e. transformation B) *is not determined* by the states of the agent's "brain". The simplest form of the learning transformation L in this case:

$$B^{t_{n+1}} = L(B^{t_n}) = B^{t_n} \otimes \xi, \tag{8}$$

where ξ is a random process (noise) and $\otimes$ denotes acting upon B.

It is obvious that the learning transformation L in the form of (8) leads not to adaptation but to degradation of the behavior. If we look at (8) as on applying a mutation ($\otimes\xi$) to the strategy of agent's behavior (B) it becomes clear that an analog of natural selection is needed to make the process adaptive.

The natural selection acts on variation in a population of individuals. The key difference of the individual learning from the evolution is that in the former the agent cannot evaluate more than one of the different variants of behavior at the same time. The solution is that during the individual learning selection acts not on the variation in the population of behaviors but on the sequence of varying behaviors. Thus the rule of individual evolutionary learning is:

"Produce blind variations of the behavior until adaptation is obtained."

Here selection is implemented as a control of blind variation. The next question, what does control variation, or who does evaluate produced behaviors? It is naturally to assume that evaluation of the behavior is performed by the agent itself. Then (8) becomes:

$$B^{t_{n+1}} = L(B^{t_n}) = B^{t_n} \otimes m(\boldsymbol{C})\xi\,, \tag{9}$$

where m is a magnitude of application of ξ to B. The value of m is determined by the state of the "brain" $\boldsymbol{C}$.

The logic of learning expressed in (9) can be summarized as follows:

1. The process of change in the behavioral strategy of the agent (learning) is not determined by the state of the agent's "brain" and has a form of blind variation of already existed behavior.
2. Amount of change through blind variation is not constant in time and is determined by the state of the agent's "brain".

Discussion

The formalization of learning as an evolutionary process presented in the previous section gives only general framework and it is insufficient for the modeling of animal learning or implementation of any learning algorithms for animats.

In the selectionist theories of learning the picture of generation of variation in the behavior seems to be straightforward. It is easily described and related to a number of mechanisms on the neuronal level such as a probabilistic pattern of connections formed during the development and spontaneous excitations of cells or assemblies. An understanding of the selection processes in the brain is a challenge.

Consider the evolution of a population consisting of agents equipped with **A** "brains". The brain **A** is a dynamical system and its dynamics can be represented as a trajectory in the phase space of possible values of the components $\boldsymbol{C}$. This trajectory determines a sequential unfolding of the agent's behavior (eq. (4)). When a new born agent has some innate behavior this behavior is represented by a primary repertoire of the trajectories. In the course of evolution the agents with adaptive sequences of actions will be selected. Hence, these innate trajectories represent the evolutionary beneficial or at least "safe" (neutral in respect to the natural selection) sequences of agent-environment interactions. Moreover, the set of innate trajectories is the only source of adaptive values for the learning process. For the behavior being adaptive these trajectories should define the target dynamical attractor. Then the goal of learning is creation of a basin of attraction for it. If the point in $\boldsymbol{C}$ which corresponds to the current state of the agent's "brain" **A** moves along the trajectory which is already "approved" by selection then no modification of the behavior (of the transformation B) is needed and $m(\boldsymbol{C}) = 0$ in eq. (9). But movement along the target trajectory might be disturbed. For example, instead of a "normal" transition $\ldots\{\boldsymbol{C}_{\boldsymbol{E}}^{t_n}, \boldsymbol{C}_{\boldsymbol{I}}^{t_n}, \boldsymbol{C}_{\boldsymbol{S}}^{t_n}, \boldsymbol{C}_{\boldsymbol{A}}^{t_n}\} \xrightarrow{B} \{\boldsymbol{C}_{\boldsymbol{E}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{I}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{S}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{A}}^{t_{n+1}}\}\ldots$ which should result in the environmental feedback $\boldsymbol{C}_{\boldsymbol{E}}^{t_{n+1}}$ the agent's "brain" might receive an "unexpected" reaction from the external world $\boldsymbol{C'}_{\boldsymbol{E}}^{t_{n+1}}$. If the "unexpected" state of the "brain" $\{\boldsymbol{C'}_{\boldsymbol{E}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{I}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{S}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{A}}^{t_{n+1}}\}$ is not lying at any part of the "safe" target trajectory then adaptation is required. To start learning the magnitude of blind variation $m(\boldsymbol{C})$ should become positive to allow generation of behavioral variations. During learning a new part of the trajectory is creating which departs from the "unexpected" state $\{\boldsymbol{C'}_{\boldsymbol{E}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{I}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{S}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{A}}^{t_{n+1}}\}$. The process of learning ends up ($m(\boldsymbol{C}) = 0$) when "approved" trajectory is reached. When after learning the agent's "brain" will sometime fall again into the state which is equivalent to $\{\boldsymbol{C'}_{\boldsymbol{E}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{I}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{S}}^{t_{n+1}}, \boldsymbol{C}_{\boldsymbol{A}}^{t_{n+1}}\}$ it will already have a trajectory to follow and no additional modifications of the behavior will be necessary.

Now, the selection criteria for the learning as evolution can be summarized as:

"The sequences of the agent-environment interactions that lead to the target trajectory should be retained".

The primary repertoire of trajectories of the agent extends only to a limited fraction of possible dimensions of the phase space. Initially only deviations along these dimensions evoke the learning process and variations in all other dimensions are "don't matter". A new branches added to the initial target trajectory by lifetime learning extend it to a new dimensions forming a secondary repertoire. Then deviations in both the primary and the secondary repertoires are used for the learning initiation and finalization.

The innate and learned behavioral trajectories are an adaptive knowledge gathered trial by trial during the evolutionary and individual history; hence, losing them is losing evolutionary advantage. This raises a requirement to the process of learning, namely, that the growth of the new branches of the attractor should not change the existing traces.

This "behavioral trajectory" analysis brings some conceptual extensions in comparison to the other selectionist approaches to the learning as evolution.

Control of learning by deviation from the target behavioral trajectory is similar to the homeostatic adaptation controlled by essential variables suggested by Ashby (Ashby, 1960). However, the homeostatic control has no mechanism for the retaining of knowledge gained during the learning through trials and errors, when the same deviation of the essential variables occurs next time the procedure of adaptation should be repeated again. Thus, Ashbian theory addresses the question of sustainability of behavior but not of its adaptive modification.

Discussing the work of Ashby Di Paolo (Di Paolo, 2003) suggested a hypothesis that is very close to the "behavioral trajectory" scheme:

> "Habits, as self-sustaining dynamic structures, underly the generation of behaviour and so it is them that are challenged when behaviour is perturbed. An interesting hypothesis is that often when adaptation occurs in the animal world this is not because organismic survival is challenged directly but because the circular process generating a habit is." (Di Paolo, 2003, p.31)

In relation to the general conceptual framework of the evolutionary epistemology (Campbell, 1974; Popper, 1984) the scheme proposed in this paper contains the next level of details. Recognition of the divergence from the target trajectory allows the agent to detect the problem situation. New behavioral trajectories sinking into the target attractor are retained, so the behavioral attractor plays the role of vicarious selector.

The selection of behavioral sequences toward attractor seems, on the first sight, to be similar to the match/mismatch selection of the neural Darwinism theories (Edelman, 1987, 1993; Changeux and Dehaene, 1989). When the agent encounters an unexpected situation it detects mismatch between perception and internal state. According to the neural selectionism the internal state is transformed to match the input. No action upon the environment is needed, neuronal dynamics only is sufficient to do that. Contrary, in the trajectory paradigm the internal state is a valuable knowledge which is kept intact and actions performed to change the situation, i.e. the input toward target values.

The analysis presented in this paper deals with the phenomenological level of description of behavior and learning. At the next level, the issue of cellular mechanisms compatible with the evolutionary scheme of learning should be addressed. The rules of cells interactions should allow detection of the deviations from the target behavioral attractor and creation of new neuronal functional systems while preserving existed ones.

Summary

The theory of learning should have an explanation why learning normally results in evolutionary adaptive modifications. The common explanatory scheme for the adaptivness of learning is based on the "feedback" logic. In this scheme the reward system of an organism or an agent evolves by the natural selection to effectively evaluate stimuli in terms of their expected contribution to the evolutionary success. The error between predicted and received reward is used as a "feedback" signal for correction of the behavior.

Variation and selection principle provides an alternative and opposite explanation to the "feedback" logic (see Table 1). Applied to the lifetime learning it assumes that at the first new behavioral variants are produced and then selected to meet evolutionary demands.

A number of approaches to utilize the evolutionary logic for the explanation of learning were proposed but the theory is still not satisfactorily. An attempt to clarify and extend the basic ideas underlying these approaches presented in this paper resulted in the following contributions:

- The innate behavior shaped by natural selection brings evolutionary values to the level of learning. This innate behavior constitutes an initial target trajectory of the agent-environment interactions.
- The generation(variation)/selection cycle of learning starts from the critical deviation from the existing behavioral trajectory and stops when deviation is eliminated.
- During learning new behaviors are created by blind variations. The behaviors leading to the target trajectory are selected.
- New branches of the behavioral trajectory produced by learning are included in the target set and start to play the role in controlling generation(variation)/selection cycle.
- The existed behavioral trajectory should not be altered as the learning unfolds. It keeps knowledge about adaptations survived selection through the evolutionary and learning history.

Acknowledgments

I deeply acknowledge the anonymous reviewers for their thoughtful suggestions and comments. Thanks to Konstantin Anokhin for the critical discussion of the ideas underlying early versions of this paper, Valentin Nepomnyashchikh and Carlos Gershenson for thier useful comments and suggestions. This work was supported by the Russian Fund for Basic Research, project 07-01-00180.

References

Adams, P. (1998). Hebb and Darwin. *Journal of theoretical Biology*, 195:419-438.

Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour*.(Second edition). London: Chapman and Hall.

Beer, R.D. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J. Meyer, J. Pollack, & S. Wilson (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, 421–429. Cambridge, MA: MIT Press.

Beer, R.D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3):91–99.

Berridge, K.C., and Robinson, T.E. (2003). Parsing reward. *Trends in Neurosciences*, 26(9):507-513.

Campbell, D.T. (1974). Evolutionary Epistemology. In P.A. Schilpp (ed.), *The Philosophy of Karl Popper*, 413–463. LaSalle, IL: Open Court.

Changeux, J.-P. Dehaene, S. (1989). Neuronal models of cognitive functions. *Cognition*, 33:63-109.

Dayan, P., and Balleine, B.W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36:285–298.

Dehaene, S., and Changeux, J. P. (2000). Reward-dependent learning in neuronal networks for planning and decision making. *Progress of Brain Research*, 126:217—229.

Di Paolo, E. A. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In: K. Murase & T. Asakura (Eds), *Dynamical Systems Approach to Embodiment and Sociality*, 19-42. Adelaide, Australia: Advanced Knowledge International.

Dorigo, M., and Colombetti, M. (1998). *Robot Shaping: An Experiment in Behavior Engineering*. Cambridge, MA: MIT Press.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Edelman, G.M. (1993). Neural Darwinism: selection and reentrant signaling in higher brain function. *Neuron*, 10:115-125.

Foundations of Science, (2001). 6 (special issue on 'Radical Constructivism and the Sciences').

Glasersfeld, E.V. (1995). *Radical Constructivism: A Way of Knowing and Learning*. London: Falmer Press.

Harvey, I., Di Paolo, E., Wood, R., Quinn, M., and Tuci, E. (2005). Evolutionary Robotics: A New Scientific Tool for Studying Cognition. *Artificial Life*, 11(1):79–98.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: The Sussex approach. *Robotics and Autonomous Systems*, 20:205–224.

Izhikevich, E.M., Gally, J.A., and Edelman, G.M. (2004). Spike-timing Dynamics of Neuronal Groups. *Cerebral Cortex*, 14:933–944.

Journal of Adaptive Behavior. (2002). 10(3-4).

Maes, P. (1994). Modeling adaptive autonomous agents. *Artificial Life*, 1(1-2):135-162.

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

Peschl, M.F. (1997). The Representational Relation Between Environmental Structures and Neural Systems: Autonomy and Environmental Dependency in Neural Knowledge Representation. *Nonlinear Dynamics, Psychology, and Life Sciences*, 1(2):99–121.

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

Popper, K.R. (1984). Evolutionary Epistemology. In J.W. Pollard (ed.), *Evolutionary Theory: Paths into the Future*, 239-255. John Wiley & Sons, Chichester and New York.

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review Neuroscience*, 23:473–500.

Skinner, B.F. (1981). Selection by Consequences. *Science*, 213(4507):501-504.

Suri, R.E., Bargas, J., and Arbib M.A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience*, 103(1):65-85.

Sutton, R.S., and Barto, A.G. (1998). *Reinforcement learning an introduction*. Cambridge, MA: MIT Press.

Toussaint, M. (2004). *The evolution of genetic representations and modular adaptation.* PhD thesis, Institut für Neuroinformatik, Ruhr-Universiät Bochum, Germany. Berlin: Logos Verlag.