Interactivism in Artificial Intelligence (AI) and Intelligent Robotics

Article *in* New Ideas in Psychology · October 2006

DOI: 10.1016/j.newideapsych.2006.07.005

Georgi Stojanov
American University of Paris

Goran Trajkovski
Western Governors University

Andrea Kulakov
Ss. Cyril and Methodius University in Skopje Macedonia

ELSEVIER

New Ideas in Psychology 24 (2006) 163–185

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

Interactivism in artificial intelligence (AI) and intelligent robotics

Georgi Stojanov[a,*], Goran Trajkovski[b], Andrea Kulakov[c]

[a] *Computer Science Department, American University of Paris, France*
[b] *Cognitive Agency and Robotics Laboratory (CARoL), Computer and Information Sciences Department, Towson University, Baltimore, MD, USA*
[c] *Computer Science Department, Faculty of Electrical Engineering, University of Sts Cyril and Methodius in Skopje, Macedonia*

Available online 11 October 2006

Abstract

This paper overviews the interactivist model of representation and its applications in artificial intelligence (AI) and intelligent robotics. Selected examples from approaches in AI and robotics are contrasted with the generic interactivist architecture in order to illustrate specific features of it. Petitagé, an artificial agent that instantiates our interactivist-expectative theory of agency and learning (IETAL), is discussed in detail from the interactivist perspective.

Keywords: Interactivism; Generic interactivist architecture; Cognitive robotics; Interactivist-expectative theory of agency and learning (IETAL); Petitagé

PsycINFO classification: 4120; 4140

1. Introduction

The central theme of this paper is the interactivist model of representation (needless to say, as understood by the authors) and its application in AI and intelligent robotics. Since our goal is to provide a broad introduction to *interactivist AI and robotics*, we first sketch a generic architecture for an intelligent agent that incorporates the key elements of the

*Corresponding author. Tel.: +33 1 40 62 06 82; fax: +33 1 45 55 17 89.
E-mail addresses: gstojanov@aup.edu (G. Stojanov), gtrajkovski@towson.edu (G. Trajkovski), kulak@etf.ukim.edu.mk (A. Kulakov).

0732-118X/$ - see front matter
doi:10.1016/j.newideapsych.2006.07.005

interactivist approach to representation. Next, we have chosen several examples of agent architectures from AI and intelligent robotics and compared them against the generic interactivist architecture, pointing out similarities and differences. Instead of attempting to give an extensive review of research approaches that have adopted (elements of) the interactivist model, we have focused on a few examples that illustrate our points.

2. Enter representation

The notion of *representation* has always been at the core of the discussions in artificial intelligence (AI), intelligent robotics, cognitive science, and other sciences of mind. Competing schools of thought have been defined according to their particular model of representation, and major changes or paradigm shifts have involved changes in the underlying model of representation. For instance, the symbolic approach takes representation to consist of symbols and their relations, which correspond to objects and relations in the environment. Connectionism understands representation as sub-symbolic and as having distributed features. Anti-representational models deny the existence of any representation whatsoever.

Since the early 1990s, we have witnessed a growing interest in theories of cognitive development of biological organisms as a potentially adequate theoretical background for understanding embodied agents. In the early 1990s, the role of embodiment in cognitive processes was being reconsidered and new ways of modeling representation were being sought. Roughly in the same period, some researchers in AI and robotics were looking to *developmental psychology* for theoretical backing; e.g., Drescher (1991), the Geneva-based CEPIAG group (Schachner, 1997), Stojanov et al. (1995, 1997, 1999, 2001) and Stojanov (1997, 1999a, b, 2001). Others were looking for inspiration in biology; e.g., Brooks (1991a, b), Dautenhahn (1998), Maes (see 1990 for an overview), Mataric (1990, 1992) and Quick et al. (1999).

The search for new conceptions of embodied agents culminated in the first Interactivist Summer Institute and the first International Workshop in Epigenetic Robotics, which were held in the summer and autumn of 2001, respectively.

3. Interactivist objections to classical symbolic representations

A number of criticisms of symbolic representation are now regarded as classic. One of these is the Frame problem (Pylyshyn, 1987), which consists of determining what pieces of the agent's vast amount of stored knowledge are relevant for the current situation or problem. The Symbol Grounding problem (Harnad, 1990) calls into question the basic notion of *transduction*, the process that supposedly transforms sensory states into symbolic representation. AI models that presume that representation is symbolic require transduction. To use Harnad's words, "How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?" The Frame-of-Reference problem (Clancey, 1989) is about the relations among the observer, the designer, the artifact, and the environment. For example, when designing a classical expert system, the knowledge engineer is at the same time an observer and a designer who observes and describes in symbolic terms the problem-solving behavior of a human expert. The categories and symbols are those of the engineer, and are grounded in his experience—not in the artifact's experience (Verschure, 1993). These problems are not merely of philosophical concern: they have manifold consequences for practice.

Bickhard's interactivism is a radical alternative to symbolic models of representation:

> [I]nteractivism is a complex philosophical and theoretical system… [that] involves a commitment to a strict naturalism. By naturalism is meant (roughly) a regulative assumption that reality is integrated; that there are no isolatable and independent grounds of reality, such as would be the case if the world were made of Cartesian substances. (2003, pp. 1–3)

Interactivism adopts *process metaphysics*: the view that the world fundamentally is organizations of processes. Bickhard (1998a) notes that in the history of the sciences, *substance-based models* have been replaced with process models as each field matures and progresses: the phlogiston model has given way to a model of the process of combustion, models of heat in terms of caloric have given way to a kinetic energy model, and *élan vital* has been superseded by self-maintaining and -reproducing organizations of processes. The only remaining exceptions to this progression are the sciences of mind (Bickhard, 1998a).

Bickhard vigorously and persuasively argues against classical representations as ill-conceived, immature *substance-based* notions (Bickhard, 1993; Bickhard & Terveen, 1995). Among other things, classical representations (often referred to as *encodings* in the parlance of interactivism) cannot account for the normativity of representations or for the emergence of new representations. Below we present several objections raised by Bickhard on various occasions.

Encodingism is circular. By postulating the existence of symbols as the basic representational units, which represent objects (actions, events, etc.) from the real world, encodingism presupposes phenomena that it is supposed to explain. Namely, it cannot explain how new symbols will be learned.

Encodingism is incoherent. Encodings are able to carry representational content when such content *is already* provided. X_n can *stand in* for (or *represent*) X_{n-1}, when X_{n-1} provides the representational content for X_n. So, all human-generated representations (maps, diagrams, drawings, etc.) convey content by virtue of pre-established conventions. They *stand in* for those conventions. The stand-in relation can be iterated many times, but there must be some bottom level X_0 of foundational encodings. There have to be representations that do not just stand in for other representations, but these representations have to be encodings, too. Bickhard calls this *the terminal incoherence*.

Encodingism relies on symbols which need to be interpreted, but it fails to account for the interpreter. Such external representations as maps, statues, and pictures are useful for their human interpreters. But it is not clear, once the maps and pictures and other encodings are assumed to be internal for some artificial system, what the interpreter is—who the interpreter is—or where it is.

What could it mean *within the cognitive system,* without direct, nonrepresentational contact with the world (whatever that would be), for the system to know that some representation is not correct? At best, if given some external feedback about the correctness of a given representation, a system that relies on symbols can replace it with (a combination of) another symbols drawn from a pre-given set.

Instead of contending with such frequently discussed issues as systematicity, productivity, and other characteristics of representational systems in linguistically capable agents, we first need to deal with the problems presented above.

4. Interactivist representations

Interactivist representations emerge out of the interaction between the agent and its environment. They are not an outcome of any sort of processing of inputs.

> Passive systems that only receive inputs *cannot* be representational in the interactive sense. No "processing" of inputs can yield interactive representationality, including any connectionist processing of inputs: representation is an emergent property of situated *interacting* systems, not passive systems. (Bickhard, 1998b)

In some of our previous works, we have put forward similar claims (Stojanov et al., 1995). This comes as no surprise because according to Bickhard (1980a, 2003) interactivism shares much with Piaget (1970) genetic epistemology—and genetic epistemology was precisely our point of departure for several projects (e.g., Stojanov et al., 1995, 1997). In our later research we explicitly embraced interactivist ideas (Stojanov, 1997, 1999a, b).

Before proceeding further with interactivist representations, we will elaborate on the generic notion of an interactivist agent.

5. Agents as action systems

Within the interactivist framework, the generic agent is seen as an action system. An action system is a system that is autonomous and stable, can perform actions in the environment it inhabits, and can sense the effects of those actions (Fig. 1).

Subsystems of the action-system are engaged in interactions with the portions of the environment (Fig. 2). The very structure (embodiment) of the agent imposes a primordial structure on the sensory influx. Certainly, there are actions internal to the agent the consequences of which can be available to the agent and which are without any immediate effect to the environment.

For example, it is impossible for, more or less, standard human agents (and without additional equipment) to see what is in front of them *and* what is behind them—at the same moment. Again, you cannot touch your chin with the thumb and the top of your head with the index finger of the same hand, simultaneously.

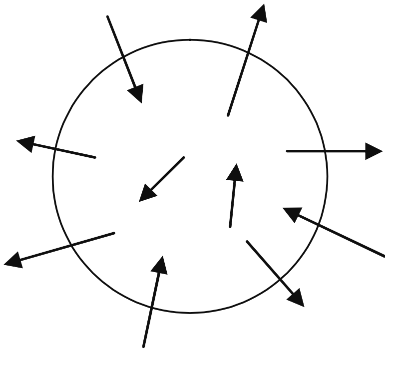

Fig. 1. Artist's rendition of an action system. The circle represents the tentative boundary of the agent; arrows represent actions. Note that even arrows coming into the system are treated as differentiating/detecting perceptual actions of the system, rather than actions of the environment on the system.

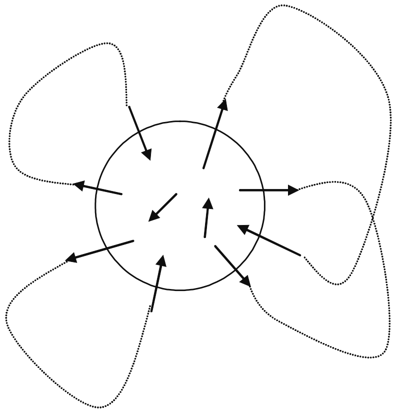

Fig. 2. Subsystems of the action system are engaged in interactions with the portions of environment.

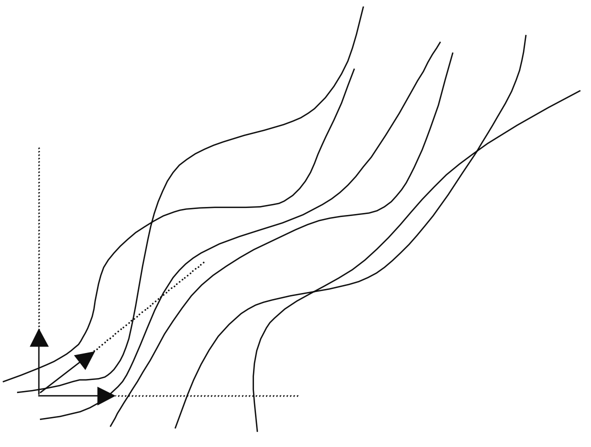

Fig. 3. Trajectories $(p_1 \ldots p_n)$ in the parameter space of an action system. One of the dimensions is time and the others depend on the nature of the parameters.

The action-system *creates* its inputs by performing various actions. The sensory-motor influx thus generated serves as a basis for interactive representation. So, we can write that

$$S = f(A,\ E),$$

where the sensory input (S) is affected by the actions (A) that the agent performs, and by the environment (E) which imposes certain constraints. Note that the notion of *action* incorporates the constraints imposed by the very embodiment of the agent (mentioned in the previous paragraph) and that to some particular agent the environment is manifested via a particular relation $f(.)$. Note also that some subsystems work *within* the action system, producing no externally observable behavior, but are still capable of inducing changes in S. In other words: the environment unfolds its structure to the agent (the action system) by putting specific constraints on how processes initiated by the agent proceed through time.

We can illustrate the dynamics of such a system by displaying the trajectories of some chosen set of parameters of the system $(p_1 \ldots p_n)$ through time (Fig. 3). One can think of

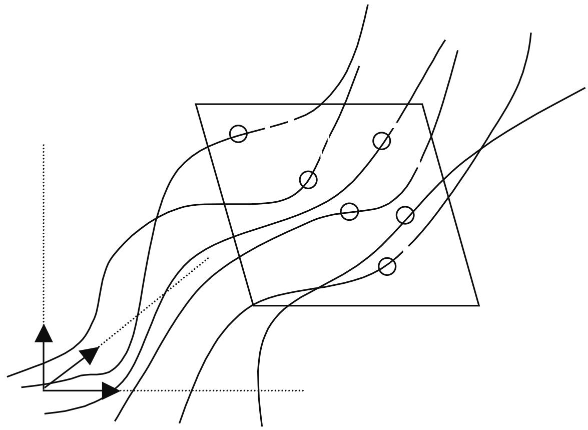

Fig. 4. The *situation image* is a snapshot of the values of parameters and the possible continuations at a given moment.

these parameters as a subset of elements of S and A; i.e., they are strictly a view from within.

At a given moment, the snapshot of the values of the parameters, as well as the indications for their possible continuations is called the *situation image* for extensive definitions of the interactivist concepts mentioned here see Bickhard, 1980b. An instant of time is represented by the plane in Fig. 4.

So far in our discussion, we have remained on a fairly general and definitional level. In the next section we resume our elaboration of interactive representation.

6. Why do agents act?

The answer to the question "Why do agents act?" is far from straightforward. An external (human) observer could describe the behavior of the agent at different levels, depending on what is taken as an element of behavior, or basic action.

Although in this context we tend to construe the verb *to act* as referring to something that the agent is doing intentionally, we would like to cover all of the processes taking place *in* and *between* the action system and its environment.

We take the *autopoietic* view (Maturana & Varela, 1980) that agents (or action systems) *act* to preserve their autonomy and to maintain the conditions for their further existence. Bickhard (1998a) often puts forward the canonical example of a candle flame. The flame maintains above threshold combustion temperature, vaporizing wax into fuel and—in standard conditions—induces convection (which brings in fresh oxygen and disperses waste products). Bickhard calls these systems *self-maintenant*. Furthermore, a system that could maintain its condition of being self-maintenant is called *recursively* self-maintenant. A candle flame certainly is not recursively self-maintenant: there is not much that the flame (or more precisely the burning candle) could do when it runs out of wax. But, a bacterium may be able to swim as long as the sugar gradient is rising, and tumble if it senses that it is swimming down the gradient (Campbell, 1990). In the above-mentioned sense, this bacterium is a primitive recursive self-maintenant system capable of switching between interactions that differentiate *good* and *bad* directions of swimming.

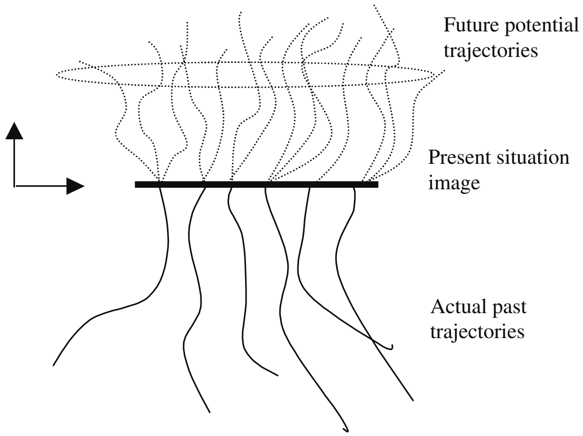

Fig. 5. A view from the above.

We see that in order to be recursively self-maintenant, an action-system has to be able to *discriminate* various aspects of the environment and evaluate them within its *inner value system*. (The values in this inner value system are different from values in the knowing-levels model. In the *knowing-levels model* (Campbell & Bickhard, 1986; Bickhard, 1998b), there are goals at Level 1 and metagoals, or values, at Level 2. The inner value system, though, is part of a larger system that functions at Level 1).

Back to Figs. 3 and 4: the situation image in a given moment should include also indicators of potentialities for interaction, which would guide the trajectories towards preferred subspaces (where the inner value system determines what is preferred). For example, an agent's representation of something that is a physically manipulable object would include indicators for the invariant patterns of interaction with that object, based on past experiences. As depicted metaphorically in Fig. 5, the situation image also includes indicators for the potential paths that may be taken. These indicators would include contextual information regarding the agent's goals, experience, and level of development.

Wrapping up this section about interactive representation we want to quote Bickhard (1980b) again:

> There is no direct or total knowledge of the world, only fallible and partial knowledge of its interactive characteristics. Thus, the world image is constructed from the specific to the general, out of the basic elements of knowledge in the *procedures innate to the system*, while the situation image is differentiated within the world image from the general to the specific by the outcomes of various interactions. The world image is a *hierarchicalized network* of general interaction possibilities and dependencies, while the situation image is a scheme structure of current possibilities. (p. 59, *emphases added*)

7. Generic interactivist architecture

Drawing on the general considerations we put forward above, and on details from our previous work (Stojanov, 1997, 1999a; Kulakov, 1998), we have proposed the essential architectural elements for an agent capable of interactive representation: structures, inner values, hierarchies, and stages (Kulakov & Stojanov, 2002).

Fig. 6 is far from being a blueprint for building an interactivist robot but it illustrates the main concepts. At the core of the system there is the agent's innate structure, which determines the actions that can be initiated by the system. There is also a *primitive innate value system* (depicted in Fig. 6 as lights with different shades of gray), which provide for the emotional coloring of the outcomes of the probes; once the agent comes into existence it starts acting in the environment according to these structures. The environment puts constraints on the way actions are being performed; a subsystem called *action-sensory flow monitor* transmits its output to a *pattern detector* subsystem which is triggered by certain patterns in the flow (see Fig. 3) and can have top-down influence on action selection, as well as generate *anticipations* about the incoming sensor inputs. All of these components are necessary so that the recursively self-maintenant action system can keep the trajectories in the "good" subspaces (cf. the bacterium example). A particular implementation would—at least—need to specify the details of the inner value system and define the pattern recognition subsystem accordingly. As the agent grows, structural changes introduce meta-pattern detectors which lead in turn to more complex value systems that

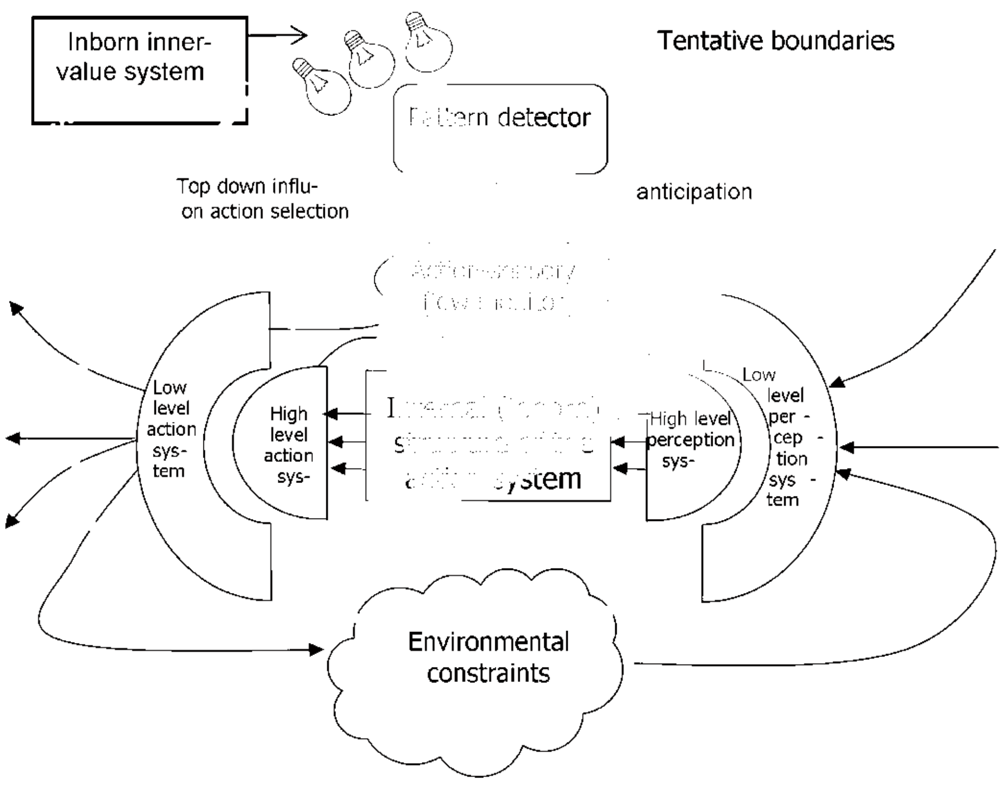

Fig. 6. Generic interactivist architecture of a "newborn" agent.

provide a basis for more complex behavior. The *tentative boundary of the agent* stresses the fuzziness of the transition between "physical" and "social" environment (De Preester, 2002), as well as the process metaphysics of the whole approach (see Fig. 7).

The next step would be to find interpretations of the concepts mentioned in the previous paragraph, in terms of *goals, motivations, emotions, cognitive development, adaptation, social and physical situatedness*, and so on.

To illustrate some of the points expressed above, we present the results of an experiment that we have performed with our existing agent Petitagé (see, e.g., Stojanov, 2001, for a detailed description of this agent).

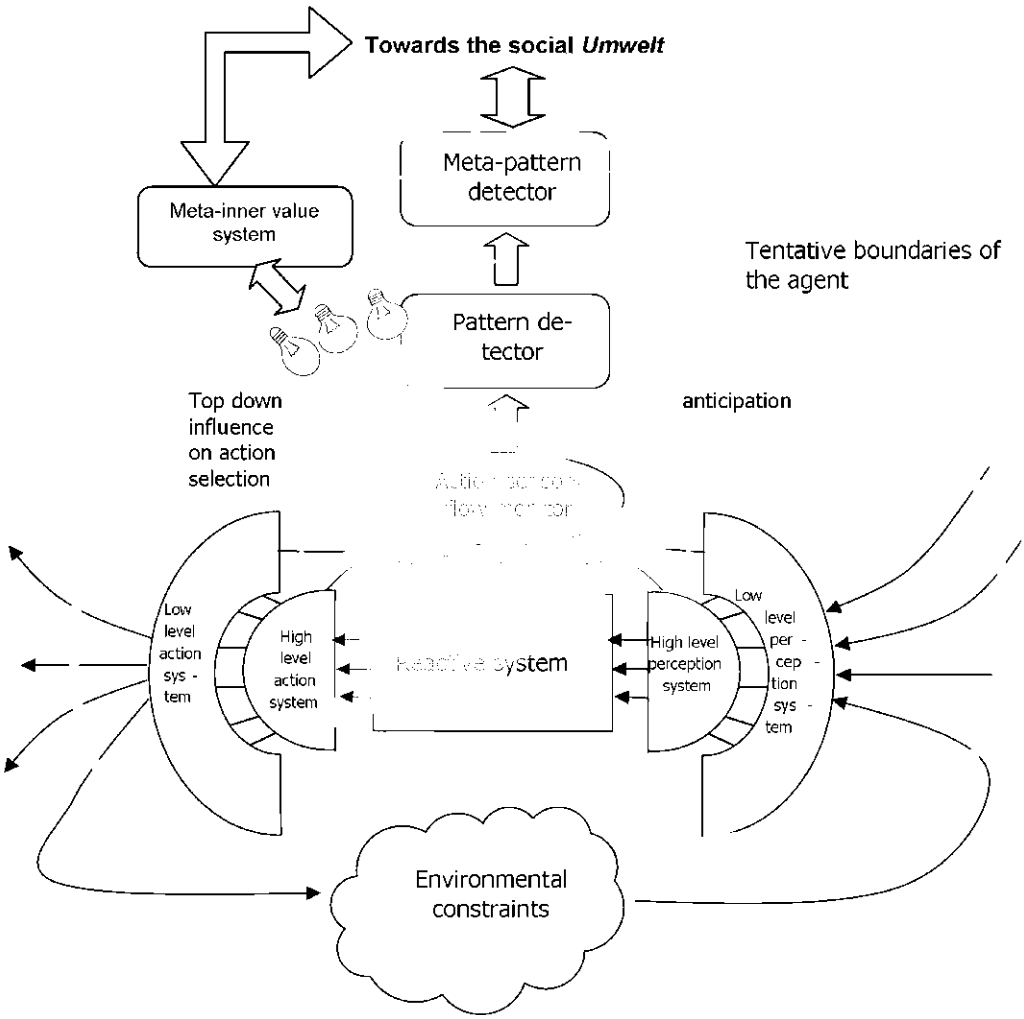

Fig. 7. A mature agent, capable of experimenting and gaining insights in its idiosyncratic physical and social environment (*Umwelt*).

8. An illustrative experiment

Originally, our Petitagé architecture was developed to illustrate the sensory-motor stage and the beginnings of the preoperational stage according to Piaget (1937/1954). By building a model of its environment Petitagé successfully solves the navigation problem in 2D maze-like environments.

Central to this architecture is the notion of an "inborn" schema. Such a schema is a particular sequence of elementary actions that the agent is capable of performing. Petitagé can perform 4 elementary actions ("f"—go forward, "b"—go backwards, "l"—go left, "r"—go right) so one possible "inborn" schema is ffrfffrfflbfff.

We have conducted an experiment with Petitagé in two different environments, starting with the same "inborn" schema. The "inborn" schema partly reflects the particularities of the embodiment of the agent.

For the sake of this experiment, we neglected the navigational system of the agent, and after choosing a particular "inborn" schema we let it cruise in two different environments. The first environment (Fig. 8) is more "corridor"-like, while the second one (Fig. 9) is more "open." Each step that the agent takes is one agent-length. The "inborn" schema at the beginning was frfrflfrflfrfrffrlfrffrflflfflffrlrlf. The agent tries to execute this "inborn" schema, but because of environmental constraints, only certain action sequences are enabled.

If the agent is in front of a "wall," and the next action to be executed is "f," then this action is skipped, until the next possible action is encountered from the "inborn" schema. In that way, in the first environment, the created schemas look like: lffffffffffffflfff, ffffffffffrfffrf, fffffffffffffffrf, frfflffffffffffffffrf, etc. while in the second environment the created schemas look like frrflfrlfrfffrlfrffrflflflfflrlf, rfrflfrrfrffrlfrrffffrrlf, frfrfl-frflffffffffllffrlrlf, frfrlfrlfrflfrlflfflffrlrlf, frrffrflfrfrffrlfrffrffffffrrlf, etc.

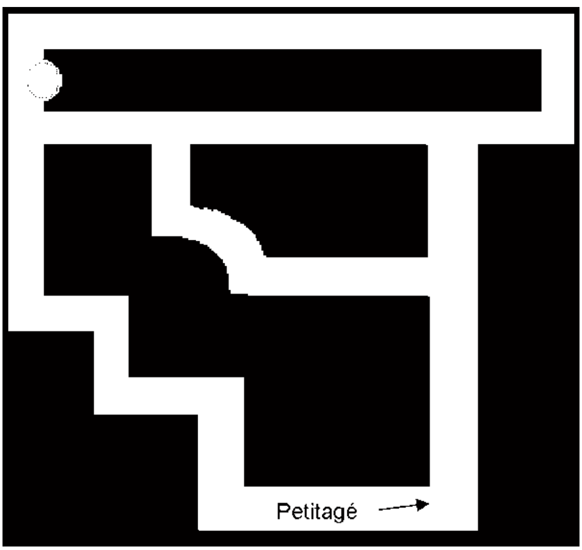

Fig. 8. Corridor-like environment.

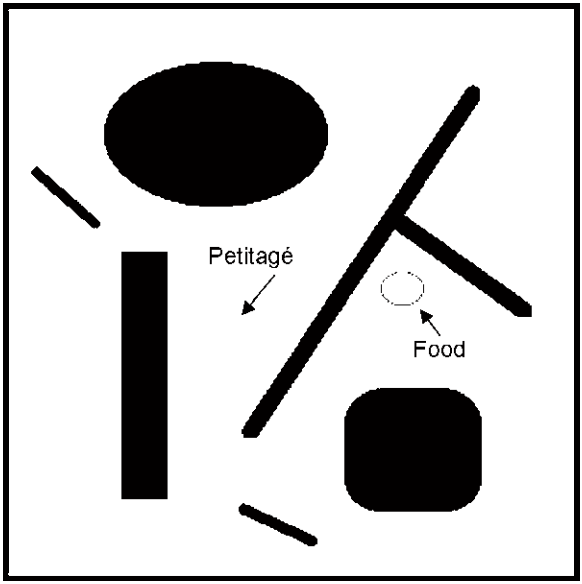

Fig. 9. An "open" environment.

It should be noted that a subsequence, as understood here, may contain actions that are not contiguous in the original schema (e.g. ab is a subsequence of accb).

It is apparent that in the corridor-like environment, the schemas contain long subsequences of "f"-s, while in the open environment the schemas more "r"-s and "l"-s than "f"-s.

Also, the average schema in the corridor-like environment (average schema length = 19.6 actions, SD = 2.18) is shorter than the average schema in the open-environment (average schema length = 26.9 actions, SD = 3.46), because of the greater constraints that the corridor-like environment imposes on the agents' actions.

What we need now is the *pattern detector* mentioned in the previous subsection, so the agent can differentiate between these two environments (and act according to its *inner value system*). Below we present one possible example of how such a pattern detector can be implemented.

In order to have a quantitative measure of the differences between the schemas in the two environments, a dissimilarity function was created that compares two sequences and after normalization returns the result in the interval [0,1], where 0 means most similar, while 1 means most different. The function that measures dissimilarity between two schemas counts all possible subsequences from the shorter schema that appear in the longer schema.

The number of all possible appearances is

$$n = \sum_{i=1}^{n_2} (n_2 - i + 1)\left(n_1 - \frac{n_2 - i}{2}\right),$$

where n_1 and n_2 are the lengths of the longer and shorter schemas respectively, which yields that

$$n = n_2(n_2 + 1)\frac{3n_1 - n_2 + 1}{6}.$$

Then the dissimilarity function is $\delta = 1-(c/n)$, where c is the number of appearances of the subsequences of the shorter schema in the longer one.

Since the dissimilarity function returns values in the upper part of the interval [0,1], we have decided to use the much smaller number

$$n = n_2\left(n_1 - \frac{n_2 - 1}{2}\right),$$

which is the number of all possible appearances of the whole shorter schema in the longer schema, while truncating values of the function δ which are smaller than zero.

The latter function is then used to produce the matrix of dissimilarities between all schemas that emerged in both environments. The matrix, in turn, serves as input to a multidimensional scaling method for grouping similar objects in increasing number of dimensions. The method gives the Kruskal's best stress value for 2 dimensions (stress: 0.38396). As can be seen on the following plot, the schemas from different environments tend to group themselves according to the environment where they were created.

The boundaries, which were drawn in Fig. 10 in freehand after the scaling, are used only to emphasize the groupings. Fig. 10 shows how different subschemas describe different environments and, consequently, different objects, that the agent enters into interaction with. One can easily imagine that objects present in the environment will be represented by particular subregions of this space according to the possibilities for interactions that they offer to the agent.

To summarize what is happening in Petitagé, we present Fig. 11. As the agent tries to exercise its inborn schemas, the environment imposes certain constraints. The mechanism that is monitoring the sensory-motor flux recognizes patterns in the flux, and at a higher level of abstraction, describes the environment in terms of these emergent patterns. The process can go to a higher level, at which meta-pattern detectors would now look for regularities in the sequences of the lower pattern detectors. In accordance with to the

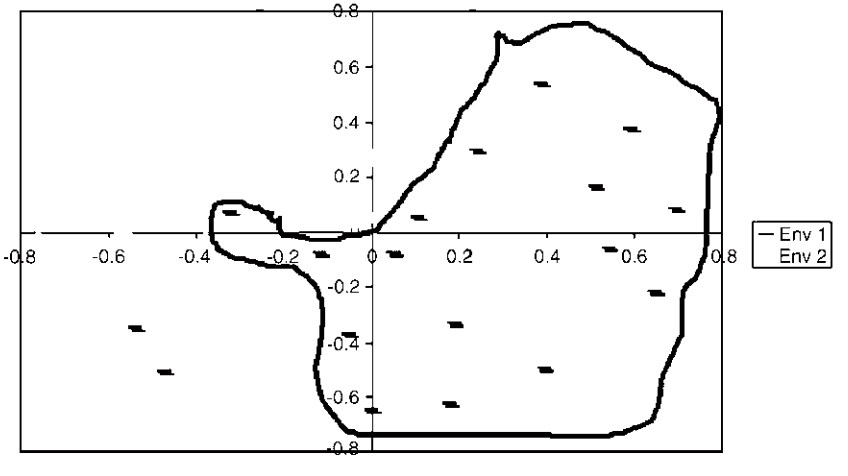

Fig. 10. Distribution of schemas in 2D after performing multidimensional scaling.

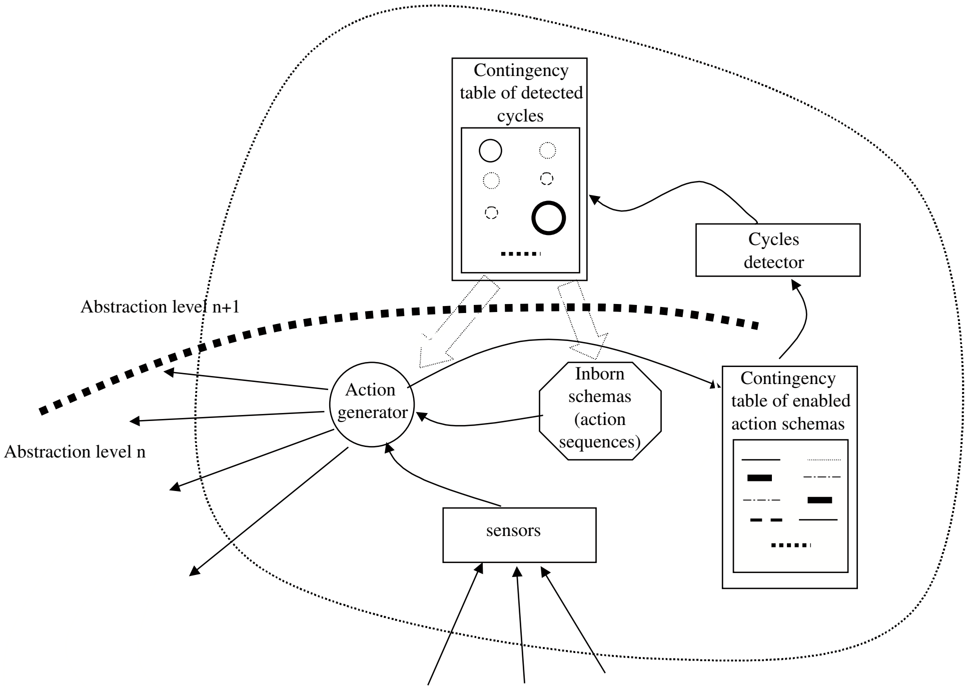

Fig. 11. Schematic description of the functioning of the Petitagé agent.

innate value system (e.g. go to the place where you'll find "food") particular trajectories are selected from the set of possible trajectories.

From what was said above, it should be clear by now what kind of properties existing agent architectures in AI and intelligent robotics will need to exhibit to support the claim that they (partly) adopt the interactivist stance. There should be at least: some *innate structure*, an *internal value system*, a mechanism which enables *recognition of emerging patterns in the sensory-motor flux* or an *abstraction mechanism*, and a way in which *these emergent structures are made causally effective* by influencing the behavior of the agent. Please note that no representation (of environment, of objects, of relations) exists in the symbolic sense. "Perceiving" objects means generating the set of potential trajectories for interaction with them.

9. AI and intelligent robotics architectures that (partly) implement interactivist representations

Gary Drescher's work (1991) is the most elaborated architecture directly inspired by Piaget's theory. From this fact alone, we should expect to find elements of the interactivist

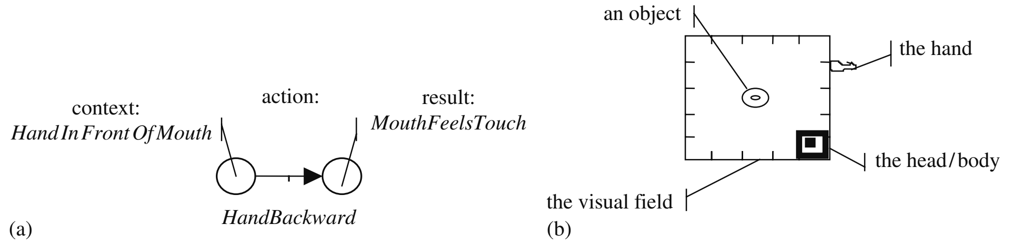

Fig. 12. (a) An example of an elementary schema for Drescher's agent; (b) the simulated world of the Drescher's agent. The visual field moves relative to the body. In this example, the agent encompasses the body and an object but not the hand.

model in it. The central place in the agent is, as expected, given to the notion of schema[1], which is formalized as the triplet *context–action–result*. Fig. 12 depicts an example of one such schema that states that if the hand is in front of the mouth, moving it backwards will result in feeling a touch on the mouth.

For Drescher, the context consists of a Boolean combination of propositions on the state of the world, which are either true or false (on or off). This world is the visual field of the agent, which moves relative to the simulated body and hand (Fig. 12b). Initially, the agent has a few schemas whose context and result fields are empty. The *schema mechanism* is equipped with a learning facility referred to as *marginal attribution*. Using this facility, the agent builds a model of its environment by learning the effects of different actions in different contexts. This is done by extensively calculating *all possible* cross-correlations (among all possible contexts, actions, and results). Through the process of *adaptation* the agent builds a meta-level of schemas based on the existing schemas, which represent the agent's Umwelt. In conclusion, Drescher's agent enters the environment with its elementary initial schemas and the structure of the environment given. By interacting with the environment, the agent establishes the consequences of its actions, and uses its schemas to achieve the goal. It does not utilize symbols in the classical AI sense (Fig. 13).

Some early work by Mataric (1990, 1992, 1995) can be viewed as highly interactivist in spirit. Her Toto robot (Fig. 14), equipped with a ring of 12 sonars and a compass, actively explores the environment in search of regions that look the same, in the sense that the perceptual input remains the same while some particular action is performed.

Thus, the environment is partitioned into regions $a_i{*}p_j{*}$ (i.e., strings of same actions concatenated with same perceptual input) and the environment model consists of the links among these regions: $link(a_i{*}p_j{*}, a_m{*}p_n{*}$—Figs. 15 and 16).

Tani (1996) describes a system (Yamabico, shown in Fig. 17) for mobile robot navigation that uses recurrent neural networks (RNNs) to generate predictions about the robot's position in a closed environmental niche. He does not offer a general framework of design principles but rather relies on the particularities of the RNN. *Attractors* that emerge in the interaction between the dynamics of the neural network and the environmental dynamics constitute the grounding of the internal symbolic description of the environment

[1]Piaget used the French word *schème*, correctly translated in English as "scheme," for elementary cognitive structures of the sort we are discussing here. However, some translations of Piaget, including Cook's (1937/1954), have rendered *schème* as "schema," and Drescher followed these sources. Although it is incorrect from a strict Piagetian standpoint, we have retained Drescher's usage to prevent further confusion.

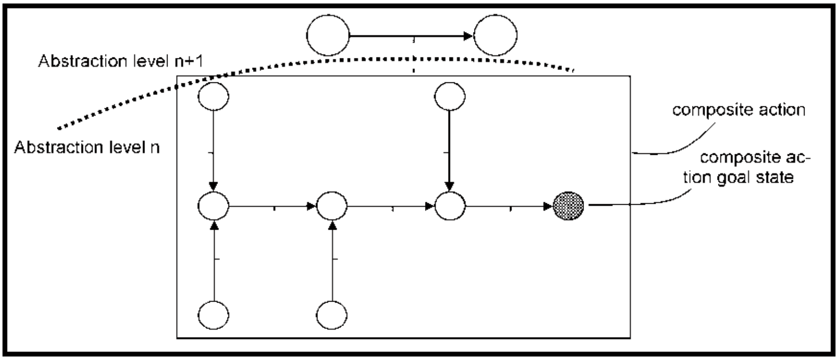

Fig. 13. Abstraction in Drescher's agent. The whole network below is subsumed in the new *more abstract action* with its own context and result nodes.

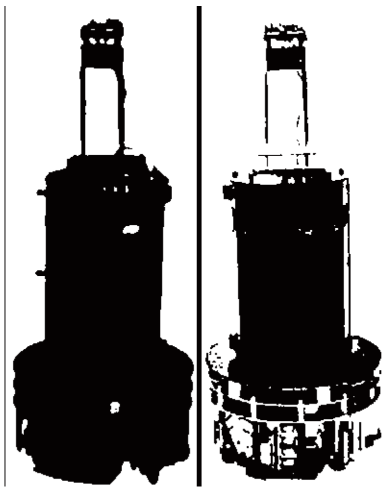

Fig. 14. Mataric's Toto Robot.

by the agent. The goal-directed planning is carried out by utilizing the acquired internal representation embedded in the intrinsic dynamics of the recurrent neural networks.

Here too, we can see how an emergent structure (the attractors) influences the behavior of the agent. The interpretation of the attractors as a higher-level abstraction of the sensory-motor flow comes naturally (Fig. 18).

Tani and Nolfi (1998), extending the main idea of Tani's previous work, introduce level structures into a system that would otherwise consist of a mixture of expert RNNs (Jacobs, Jordan, Nowlan, & Hilton, 1991). They consider multiple levels of modular networks. Now, when a particular place in the environment is recognized, the output of a lower level RNN (say, activation of some output node or nodes is given to the RNN of the level above). When the system repeatedly experiences a similar sequence of RNN module activation in the lower level, this sequence can itself be learned by a RNN module in the higher level. Each module at the higher level learns to encode a different sequence of the primitives, in terms of RNN module activations, that is reusable in different situations.

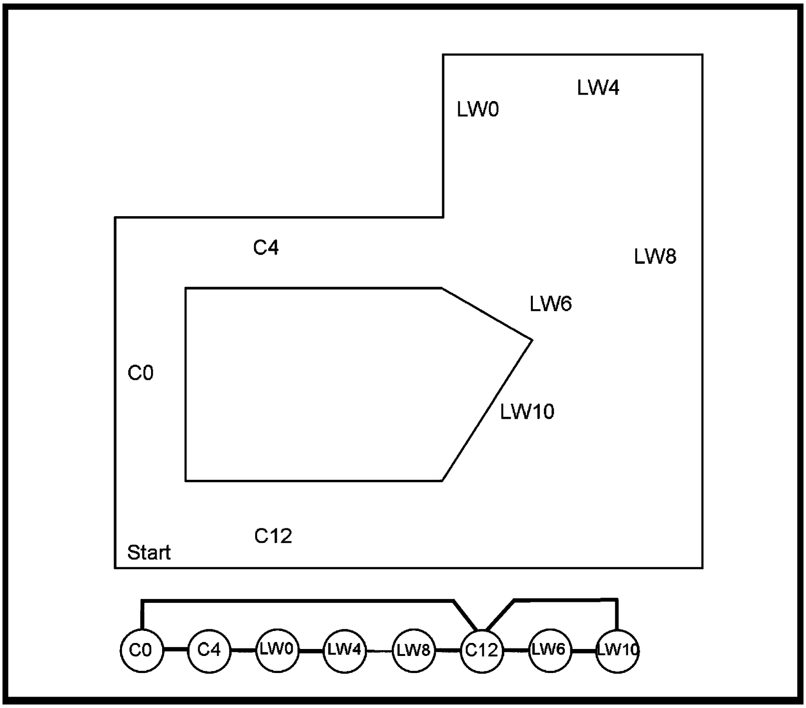

Fig. 15. Toto's representation of one floor of the MIT AI department.

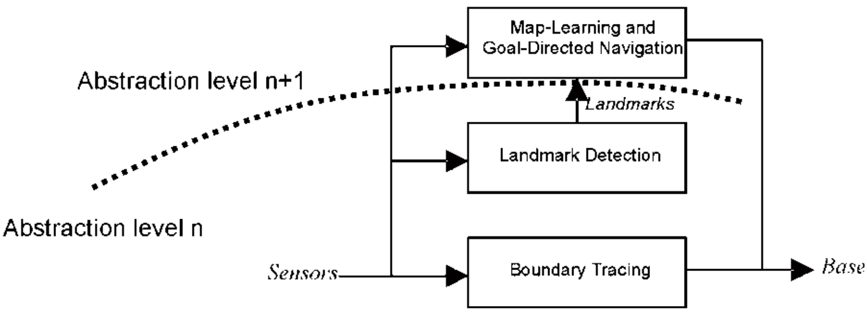

Fig. 16. Abstraction in Toto; the upper level generates expectations for sensory input while moving the robot toward the goal place.

In Fig. 19 we can see the basic RNN modules (b, c), and a part of the overall system showing two levels of the hierarchy (a).

A nice example of the abstraction process in an agent with this architecture can be seen when it is put in the kind of environment shown in Fig. 20. After spending some time in "room" A the first layer learns "primitive" concepts like *corridors*, *corners*, and *crossings*. The second level then learns to distinguish between room A and room B by relying on

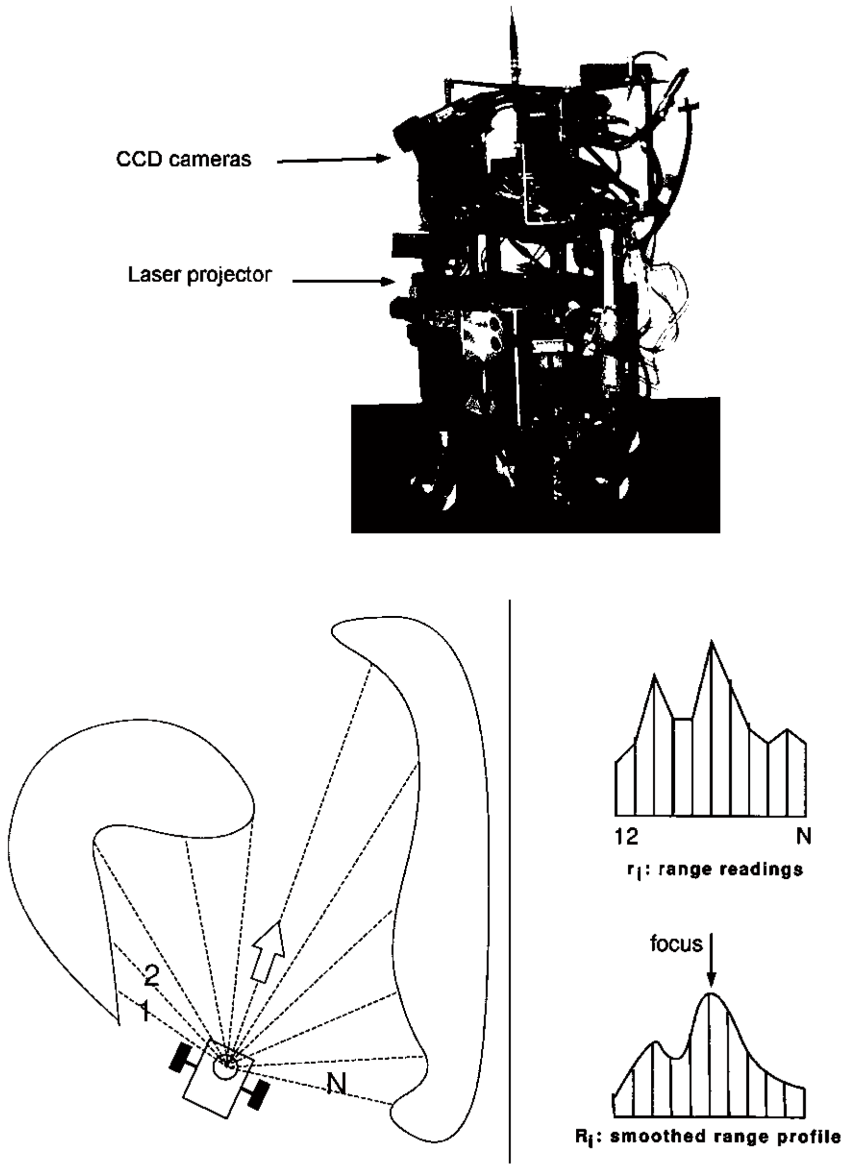

Fig. 17. Tani's Yamabico mobile robot and schematic representation of its sensory input.

regularities in the sequences of primitive concepts, so that we may think of it as the agent has learned the concept of a specific room.

The learning mechanism of Cohen's et al. (1996) simulated agent Neo is very similar to Drescher's marginal attribution. The Neo project built a virtual infant that learned many of the cognitive skills that are expected from a 3-year-old. A conceptual structure that identifies classes, and supports judgments of similarity underlies these skills. Neo's *fluents* represent objects, states, dependencies and activities. Cohen and his co-investigators were

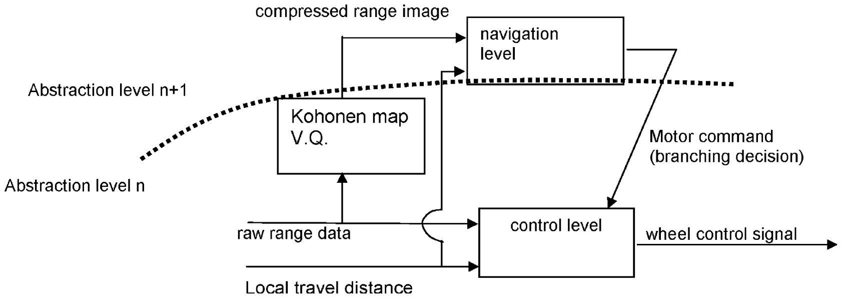

Fig. 18. Abstraction in Yamabico; the navigation level generates actions and expectations for sensory input while moving the robot toward the goal place.

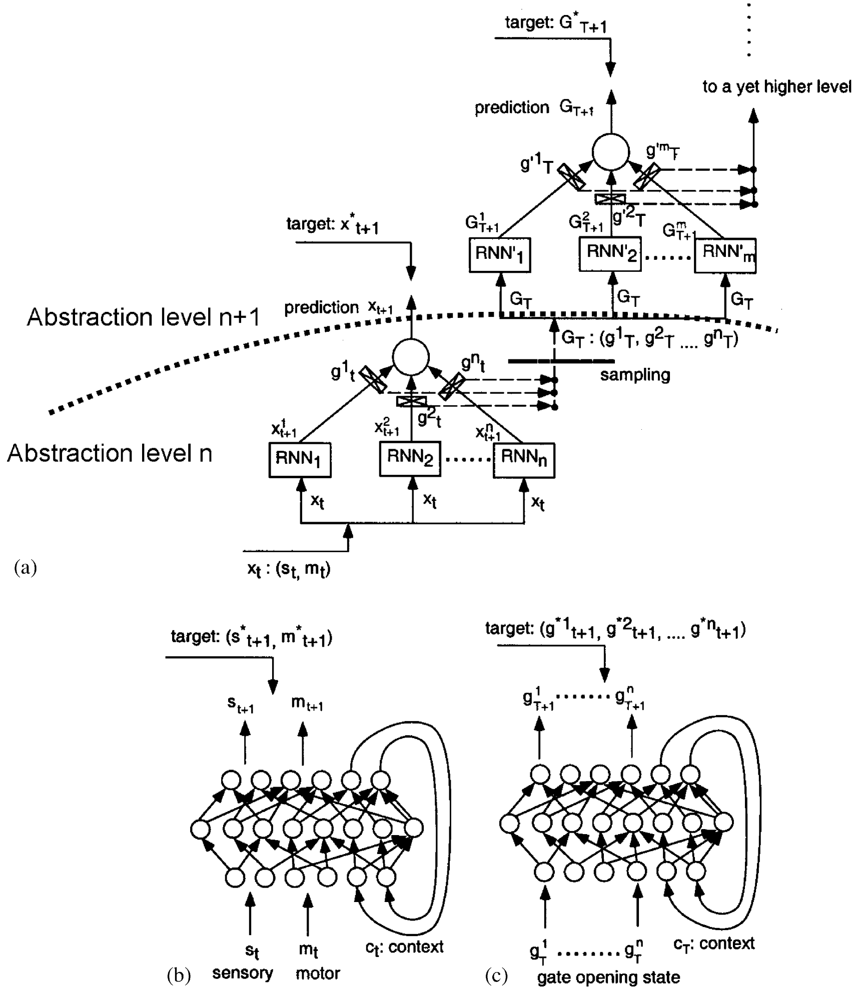

Fig. 19. Basic RNN modules (b, c), and a part of Tani and Nolfi's overall system showing two levels of the hierarchy (a).

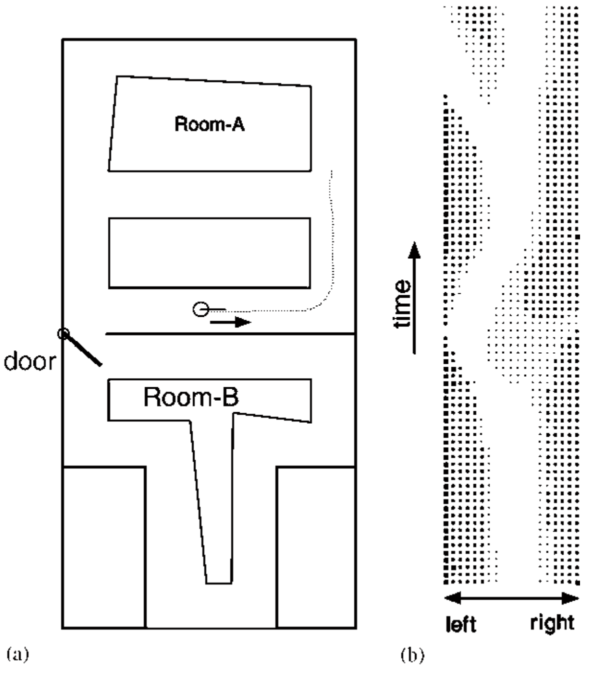

Fig. 20. The environment of Tani and Nolfi's agent. The agent learns the "primitive" concepts of corners, corridors, and crossings out of the regularities in the sensory-motor flux; at the second level, it learns the notions of room A and room B, from the regularities in the occurrences of these primitive concepts.

able to identify classes by examining Neo's learned activities, thus providing evidence that conceptual structures can be learned by interacting with the environment. The agent Neo walks through the environment so that it can move its arm through the simulated space (i.e., vary its joint angles through configuration space) in an exhaustive fashion, at a given level of resolution. When the agent encounters an object, its reflexes may cause it to interact with the object, depending on the details of the encounter.

Cherian and Troxell (1995) propose a functional model of representation for behavior-based systems. They note that

> [A]lthough there are many behavior-based robot control practitioners, the nature of representations in behavior based systems has not been investigated in detail except to show that explicit correspondence based representations are not necessary for building competent autonomous systems. (1995, p. 692)

Cherian and Troxell's *behavior networks* consist of a collection of task-achieving behaviors that interact amongst themselves and with the environment, thus generating a global behavior that promotes the agent's survival while it carries out its tasks. *Representations* and *motivations* in their system are seen as *aspects* of the underlying control structures, instead of *elements* out of which it is built. The authors explicitly appeal to Bickhard's interactivism. They rightly point out that:

> Within the behavior network architecture, activation and excitation status of behavior modules and the activity levels along excitatory and inhibitory connections

> between behaviors, constitute indications of interaction potentialities that are anticipatory in nature. (p. 694)

But in their theory they do not provide mechanisms for *learning* those connections among the behavior modules. The topology is hand coded—which is not in the spirit of Bickhard's theory where the *indications of interaction potentialities* are to be learned by the agent. As a matter of fact, a sufficiently primitive interactive agent could be strictly a knowing system from an interactive point of view, but the goals in Cherian and Troxell's system require an agent capable of learning.

In their paper Parisi and Cecconi (1995) make a distinction between two different modes of learning in neural networks. The first is the traditional, passive mode, when the networks incorporate in their internal structures the regularities present in the input and the teaching input that they passively receive. In their critique of connectionist modeling, Bickhard and Terveen (1995) assume the passive model. In a second mode, which Parisi and Cecconi call *active,* the network is equipped with a repertoire of actions and can partially control its input. Networks that function in the active mode are called *ecological networks* because they are seen as agents living in some environment where they learn to predict the consequences of their actions. They then empirically show that "hidden" environmental properties are better learned in active mode. Parisi and Cecconi provide an example of how adding simple control over perception—that is, the capacity for autonomous interaction—enables the construction of better representations. To some extent, Parisi and Cecconi depart from classical connectionism by adding top-down control signals over the network input.

In a similar vein, Rao and Ballard (1995, 1996) propose architectures for modeling the capacity for invariant recognition, motion, and stereo vision. The key point is that, apart from the input-driven bottom-up signals, in their model they include *expectation-driven* (or top-down) signals. The two types of signals are dynamically combined in order to predict the current recognition state. The model is based on a hierarchical form of extended Kalman filters, according to the minimal description length principle (MDL). In particular, this model successfully predicts the properties of the neural response in the visual cortex during visual recognition and fixation tasks. The key point is that perception is construed as an active process of constructing the expected input, rather than a passive input processing.

As a last example, we want to present a model of rhythm recognition and perception intended as a proof of concept for the interactivist position. Buisson (2004) has constructed a fairly simple Java program called Tam-Tam that implements the most important ideas of interactivist representation as well as assimilation schemes à la Piaget. The stress in the model is given to the inherent temporal dimension of perception as well as its mutual determination by action. The user simply strikes a single key on the keyboard, according to certain rhythmic pattern. The program is supposed to predict the next stroke; i.e., to *learn* the rhythm. This is done without any explicit recording of the input flow. Instead, the program manages a set of sensory-motor schemes (SMS) which are analogous to Piaget's assimilating schemes. These SMSs react to the actions from the world (the keystrokes) and accommodate to these actions in order to maximize the assimilation. Though not going into the detailed description of the program, we will nevertheless try to explain its functioning. In the very beginning there is a single simple SMS: a half-tone which corresponds to a simple monotonic tam-tam. The program tries to assimilate everything which comes from the world using this SMS. Obviously, unless the user taps this simple tam-tam rhythm the program will not be very successful. When the user starts to use more complex patterns of strikes, the program has two possibilities to account for them: stretching—or compressing—the intervals between the strikes (which actually represents fine tuning of some SMS which is otherwise well suited to the input) or, in case this does not bring satisfactory stroke prediction, the program generates new SMSs by a variation/selection algorithm. Thus, the population of SMSs evolves as new SMSs are created via substitution, deletion, or insertion of notes. Every new SMS is then evaluated according its success of predicting the next stroke from the user; i.e. how well it assimilates the current rhythm. Each SMS is in fact a program thread running in parallel with other SMSs, indicating potential rhythms which are more or less in accordance with the user input. The one which best predicts the timing of the upcoming stroke is chosen to generate the anticipation of the next time interval. This simple program (only about 400 lines of JAVA code) illustrates in an elegant manner many features of the interactivist model: the inherent temporality of interaction and participation; the emergence of representation out of interactive anticipatory organization; the anticipation of truth value in anticipations, using internally detectable representational error; the inherent variation and selection constructivist mechanism of learning, which is controlled by this internal error detection.

10. Conclusion

The interactivist model in its entirety can be used to explain phenomena in various disciplines and on various levels. Here, we focused our discussion on the problem of representation and on interactivist-flavored solutions to it in AI and intelligent robotics. In conclusion, we can say that there is apparently not yet a full-fledged implementation of Bickhard's interactivist model in these fields. Neither it seems that one can be expected in the near future. A main shortcoming of each model presented here is that the number of abstraction levels and the criterion for their introduction are hand-coded. This, in our opinion, will remain a significant challenge for future implementations of the interactivist model.

Acknowledgments

This project has been funded in part by the National Academy of Sciences (NAS), under the Twinning Program, supported by contract No. INT-0002341 from the National Science Foundation (NSF). The contents of this publication do not necessary reflect the views or policies of NAS or NSF, nor does mention of trade names, commercial products, or organizations imply any endorsement by NAS or NSF.

The present article is partly based on the tutorial "Interactivism in AI," presented by Stojanov and Trajkovski at the First Interactivist Summer Institute at Lehigh University, Bethlehem, PA, in July 2001, and partly on the paper by Kulakov and Stojanov (2003). The authors are grateful to Robert Campbell and Mark Bickhard for their constructive and extremely helpful comments and suggestions on earlier drafts of this paper.

Dr. Stojanov wishes to express his gratitude for support from the *Fondation Archives Jean Piaget* in Geneva, which enabled a 3-month stay during 2004 at the Archives and provided a stimulating working environment. He is also very grateful to Prof. Jacques

Vonèche for engaging discussions, and for introducing him to the members of the CEPIAG group.

Dr. Trajkovski's work on this paper was also partly supported by the Faculty Research and Development Committee funds of Towson University, Towson, MD.

References

Brooks, R. A. (1991a). How to build complete creatures rather than isolated cognitive simulators. In Kurt van Lehn (Ed.), *Architectures for intelligence: The twenty-second carnegie mellon symposium on cognition* (pp. 225–240). Hillsdale, NJ: Erlbaum.

Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Bickhard, M. H. (1980a). A model of developmental and psychological processes. *Genetic Psychology Monographs*, *102*, 61–116.

Bickhard, M. H. (1980b). *Cognition, convention, and communication*. New York: Praeger.

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, 285–333.

Bickhard, M.H. (1998a). *Process and emergence: Normative function and representation.* Departments of Philosophy and Psychology, Lehigh University, Bethlehem, PA. URL: 〈http://www.lehigh.edu/~mhb0/ProcessEmergence.pdf〉.

Bickhard, M. H. (1998b). Levels of representationality. *Journal of Experimental and Theoretical Artificial Intelligence*, *10*(2), 179–215.

Bickhard, M.H. (2003). *Interactivism: A manifesto*. Departments of Philosophy and Psychology, Lehigh University, Bethlehem, PA. URL: 〈http://www.lehigh.edu/~mhb0/InteractivismManifesto.pdf〉.

Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in Artificial Intelligence and Cognitive Science: Impasse and solution*. Amsterdam: North-Holland.

Buisson, J.-C. (2004). A rhythm recognition computer program to advocate interactivist perception. *Cognitive Science*, *29*, 75–88.

Campbell, D. T. (1990). Levels of organization, downward causation, and the selection-theory approach to evolutionary epistemology. In G. Greenberg, & E. Tobach (Eds.), *Theories of the evolution of knowing* (p. 117). Hillsdale, NJ: Erlbaum.

Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages*. Basel: Karger.

Cherian, S., & Troxell, W. O. (1995). Interactivism: A functional model of representation for behavior-based systems. In F. Morín, A. Moreno, J. J. Merelo, & P. Chacón (Eds.), *Advances in artificial life: Proceedings of the third European conference on artificial life, Granada, Spain* (pp. 691–703). Berlin: Springer.

Clancy, W. J. (1989). The frame of reference problem in the design of intelligent machines. In van Lehn, K. (ed.) *Architectures for intelligence: The twenty-second Carnegie symposium on cognition* (pp. 357–423). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, P.R., Oates, T., & Atkin, M.S. (1996). Preliminary evidence that conceptual structure can be learned by interacting with an environment. *AAAI fall symposium on embodied cognition and action*, TR FS-96-02.

De Preester, H. (2002). Intentionality and the inside/outside distinction in sensitive systems. *Consciousness & Emotion*, *3*(1), 65–79.

Drescher, G. L. (1991). *Made-up minds: A constructivist approach to artificial intelligence*. Cambridge, MA: MIT Press.

Dautenhahn, K. (1998). The art of designing socially intelligent agents—Science, fiction, and the human in the loop. *Applied Artificial Intelligence Journal*, *12*(7–8), 573–617.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*, 335–346.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*(1), 79–87.

Kulakov, A. (1998). *Vygovorotsky—A model of an anticipative and analogy-making actor.* Unpublished M.Sc. Thesis, New Bulgarian University, Sofia.

Kulakov, A., & Stojanov, G. (2002). Structure, inner values, hierarchies, and stages: Essentials for developmental robot architecture. In *Second international workshop on epigenetic robotics*, *Lund University Cognitive Studies*, vol. 94 (pp. 63–70).

Maes, P. (Ed.). (1990). *Designing autonomous agents: Theory and practice from biology to engineering and back*. Cambridge, MA: MIT Press.

Mataric, M. (1990). Navigating with a rat brain: A neurobiologically-inspired model for robot spatial representation. In J.-A. Meyer, & S. Wilson (Eds.), *From animals to animats: International conference on simulation of adaptive behavior* (pp. 169–175). Cambridge, MA: MIT Press.

Mataric, M. (1992). Integration of representation into goal-driven behavior-based robots. *IEEE Transactions on Robotics and Automation*, *8*(3), 304–312.

Mataric, M. (1995). Integration of representation into goal-driven behavior-based robots. In L. Steels, & R. A. Brooks (Eds.), *The artificial life route to artificial intelligence* (pp. 165–186). Hillsdale, NJ: Erlbaum.

Maturana, H., & Varela, F. J. (1980). *Autopoiesis and cognition*. Dordrecht: D. Reidel.

Parisi, D., & Cecconi, F. (1995). *Learning in the active mode. Lecture Notes in Computer Science*, vol. 929. In *Proceedings of the third European conference on advances in artificial life* (pp. 439–462). New York: Springer.

Piaget, J. (1954). *The construction of reality in the child. (M. Cook, Trans.)*. New York: Basic Books (Original work published 1937).

Piaget, J. (1970). *Genetic epistemology*. New York: Columbia University Press.

Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Quick, T., Dautenhahn, K., Nehaniv, C.L., & Roberts, G. (1999). On bots and bacteria: Ontology independent embodiment. In *Proceedings of the fifth European conference on artificial life* (ECAL'99), Lausanne, Switzerland. URL: 〈http://citeseer.ist.psu.edu/quick99bots.html〉.

Rao, R., & Ballard, D. (1995). Learning saccadic eye movements using multiscale spatial filters. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in neural information processing systems, vol. 7* (pp. 893–900). Cambridge, MA: MIT Press.

Rao, R., & Ballard, D., (1996). *A class of stochastic models for invariant recognition, motion, and stereo*. Technical Report 96.1, National Resource Laboratory for the Study of Brain and Behavior, University of Rochester.

Schachner, W. (Ed.). (1997). Technical report of the constructivism, interactionism and their applications CEPIAG seminar organized in Geneva for the Growing Mind Conference, 1997.

Stojanov, G. (1997). *Expectancy theory and interpretation of EXG curves in context of biological and machine intelligence.* Unpublished doctoral dissertation, ETF, Sts. Cyril and Methodius University, Skopje, Macedonia.

Stojanov, G. (1999a). *A model for spatial representation with controlled partitioning of the environment*. Technical Report of the Workshop on Learning and Perception in Robot Systems, Parma, Italy.

Stojanov, G. (1999b). Embodiment as metaphor: Metaphorizing in the environment. In C. Nehaniv (Ed.), *Lecture Notes in Artificial Intelligence*, vol. 1562 (pp. 88–101). New York: Springer.

Stojanov, G. (2001). Petitagé: A case study in developmental robotics. In *Proceedings of the first international workshop on epigenetic robotics*, Lund University Cognitive Studies, vol. 85. URL: 〈http://www.lucs.lu.se/epigenetic-robotics/Papers/Stojanov.pdf〉.

Stojanov, G., Bozinovski, S., & Trajkovski, G. (1997). Interactionist-expectative view on agency and learning. *IMACS Journal for Mathematics and Computers in Simulation*, *44*(3), 295–310.

Stojanov, G., Kulakov, A., & Trajkovski, G. (1999). Investigating perception in humans inhabiting simple virtual environments: An enactivist view. Published abstracts of the *cognitive science conference on perception, consciousness and art*, Vrije Universitaet, Bruxelles.

Stojanov, G., Stefanovski, S., & Bozinovski, S. (1995). Expectancy based emergent environment models for autonomous agents. In *Proceedings of the fifth international symposium on automatic control and computer science*, Iasi, Romania.

Stojanov, G., & Trajkovski, G., (2001). *Interactivism in AI.* Tutorial presented at the First Interactivist Summer Institute, Lehigh University, Bethlehem, PA. URL: 〈http://www.lehigh.edu/~interact/isi2001/tutorials.html#geos〉.

Tani, J. (1996). Model-based learning for mobile robot navigation from dynamical system perspective. *IEEE Transaction on Systems, Man and Cybernetics Part B*, *26*(3), 421–436.

Tani, J., & Nolfi, S. (1998). Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. In R. Pfeifer, B. Blumberg, J.-A. Meyer, & S. Wilson (Eds.), *From animals to animats 5: Proceedings of the fifth international conference on simulation of adaptive behavior*. Cambridge, MA: MIT Press.

Verschure, P. (1993). Formal minds and biological brains: AI and Edelman's extended theory of neuronal group selection. *IEEE Expert*, *8*(5), 66–75.