PERGAMON

Neural Networks 11 (1998) 1531–1549

Neural Networks

1998 Special Issue

A bottom up approach towards the acquisition and expression of sequential representations applied to a behaving real-world device: Distributed Adaptive Control III

Paul F.M.J. Verschure[a,*], Thomas Voegtlin[b]

[a]Institute of Neuroinformatics, ETH-UZ, Gloriastrasse 32, CH-8006 Zürich, Switzerland
[b]Ecole normale supérieure, Lyon, France

Received 6 November 1997; revised 5 March 1998; accepted 5 March 1998

Abstract

Biological systems display a high degree of flexibility in problem solving. In this paper a model is presented, Distributed Adaptive Control III (DACIII), which is aimed at understanding these forms of behavior. DACIII is part of a larger modeling series directed at understanding how biological systems acquire, retain, and express knowledge of the world. This modeling series has its roots, on one hand, in the methodological consideration that brain and behavior need to be modeled from a multi-level perspective. On the other, the importance of the acquisition of representations of events in the world, as opposed to an a priori specification, is emphasized. DACIII is presented against the background of the paradigms of classical and operant conditioning. On the basis of an analysis of these experimental approaches towards the study of animal behavior a theoretical framework is defined aimed at identifying the minimal requirements of a control structure which could display these behaviors. The proposed model is evaluated in different configurations using both simulated and real robots. It is demonstrated that DACIII is able to fully bootstrap itself from a mode of control which solely relies on proximal sensors and predefined reflexes, to a level of control which is dominated by acquired representations of events transduced by distal sensors. This transition is reflected in the performance of the behaving device, from strongly variable trajectories to highly structured behavioral sequences. The results are compared with alternative models of classical and operant conditioning and models which attempt to capture the properties of its underlying neural substrate. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Conditioning; Learning; Robot; Model; Brain; Sequential representations; Problem solving; Goal directed behavior; Distributed Adaptive Control; Foraging

1. Introduction

Biological systems demonstrate a high degree of robustness in the face of environmental uncertainty. For instance, a rat placed in a seven-arm maze, each arm containing a number of food items, will rather quickly display a performance which is described as an optimal strategy (Roberts, 1992). Optimality, in this case, has an operational definition in terms of the relationship between the distance traveled and the number of food items recovered. Dependent on the task demands, for instance defined by the effort required to recover the food items, a different behavioral strategy is displayed. In case the food dispensers are covered, the animal will after training, first visit those dispensers which contain the maximal number of food pellets. In case the food is readily accessible a so-called linear strategy is followed where the nearest dispensers are visited first. Hence, dependent on the properties of the task and the environment the animal displays a different behavior; in both cases, however, converging to an optimal strategy. This type of performance relies on the balancing of many different components. For instance, the actual data available to the animal is only presented to it in egocentric coordinates. Only through defining the temporal relationships of the local 'views' of the world, together with the displayed local actions, can global 'world centered' relationships be defined. In contrast, most robotic applications dealing with issues of path planning, for instance, solely rely on global information regarding the environment (see Kröse and Van Dam, 1997, for a review). Biological systems unfortunately do not have this luxury. In addition, only a small fraction of the actual impressions of the world transduced through the sensors pertains to the task at hand. The task being defined

* Corresponding author. Tel: 0041 1 634 2680; Fax: 0041 1 634 4983; e-mail: pfmjv@ini.phys.ethz.ch

PII: S0893-6080(98)00029-X

in terms of the ‘goals’ of the animal, for instance foraging for food in case it is food deprived, and the relevant reinforcement encountered in the world. In these terms, even a seemingly straight forward behavior turns out to be a feat of problem solving. The modeling study presented in this paper is aimed at understanding the processes involved in acquiring and expressing these forms of behavior. Preliminary results of this study have been presented by Verschure (1993a).

Moore (1956) showed that it is in principle impossible to decide between alternative functional models of an observed response function. In practice, this problem of indeterminancy is often encountered. For instance, in the explanation of the types of behavior displayed in the foraging task, which can be seen as a form of operant, or instrumental, conditioning, a large range of models have been proposed. On one extreme there is the strict stimulus response interpretation which goes back to Thorndike’s law of effect (Thorndike, 1911). The law of effect specifies that in case a response leads to a ‘satisfying state of affairs’ it is ‘stamped in’ while if it leads to an ‘annoying’ state of affairs it will be ‘stamped out’. This proposal attempts to explain changes in behavior owing to conditioning solely in terms of the effects of particular actions. It has formed a center piece of the extreme behaviorist movement of Watson and Skinner. Other proposals, however, emphasized the role of the expectations the learning system entertains, for instance by Hull (1943) (see Mackintosh, 1972 and Dickinson, 1994 for a review). In this proto-cognitive approach, variables internal to the organism were introduced in the explanation of learning phenomena. One problem underlying this dilemma in theories of learning is that both the observations to define and to validate these proposals are derived from the same level of description, in this case behavior. In order to solve this problem of indeterminancy a method of convergent validation was introduced (Verschure, 1997a) which proposes that in order to enhance the probability that a model provides a unique formulation of a phenomenon it needs to be validated against constraints derived from multiple levels of description. In our present exploration, these levels are provided by the behavioral and the neuronal perspectives. The above methodological consideration provides a strong argument for a synthetic research program, which relies on large scale computer simulations interfaced to real-world devices. This seems the most effective way to actually develop and validate these ‘multi-leveled’ scenarios. The choice in the presented study to validate the model using simulated and real robots is an implication of the method of convergent validation, next to the observation that behavior can only be explained as a real-world real-time phenomenon. Verschure (1993b, 1997a) further elaborates on the methodological and conceptual arguments for this choice.

The present modeling study is part of a larger series, called Distributed Adaptive Control (DAC) (Verschure et al., 1992). The focus of these efforts are the study of the acquisition, retention, and expression of knowledge by biological systems. Part of its theoretical considerations were derived from the observed limitations in the program of artificial intelligence and some of its more recent incarnations, connectionism, new artificial intelligence, and artificial life which have been extensively analyzed over the last years (Verschure, 1990, 1992, 1993b, 1996). The bottom line of this analysis is that even though the metaphors can be changed from the digital computer to the brain in most cases the hard problem of a prioris remains; how can we explain or create adaptive behavior without assuming it beforehand? The combination of both the methodological considerations, regarding the validity of our scientific efforts, and the theoretical ones, addressing the genesis of knowledge in biological systems, constitutes a program of synthetic epistemology (Verschure, 1998).

In the present proposal, we make the assumption that in order to explain the forms of learning expressed in, for instance, a foraging task, three strongly coupled levels of control need to be distinguished (Fig. 1). First, by solely relying on prewired reflexive relationships between sensory events and actions the system functions as a reactive controller. It will reflexively respond to immediate events. Second, as an adaptive controller the system will develop representations of events that correlate in some way with stimuli which triggered the reflexes. In addition the reflexive actions can be reshaped in order to better reflect the properties of the environmental perturbation. At the level of reflective control more extended representations of sensory events and motor actions will be formed, for instance expressing their relationship over time. The behavior displayed is influenced by internally generated expectations of the properties of the world. A system which comprises of these three components will be referred to as a complete learning system. The three levels of control will generate distinct behavioral signatures. Ranging from the strongly variable behaviors displayed by a reactive control system to the highly structured behavioral patterns generated through reflective control. The goal of our modeling efforts now becomes the study of the complete learning system.

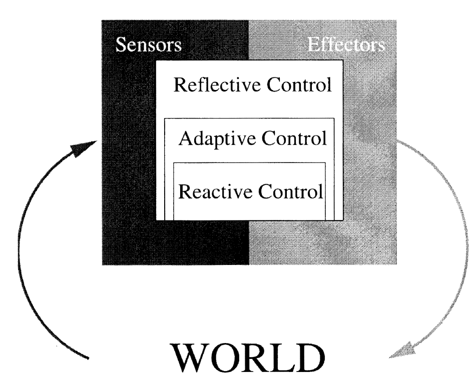

Fig. 1. The three levels of control.

2. Methods

2.1. Terminology

The study of learning and problem solving has been systematically pursued for the last century. The main paradigms that have been developed are those of classical and operant conditioning. The models presented in this study take their terminology from these domains and will be shortly described.

Classical conditioning (Pavlov, 1927) refers to learning phenomena where initially neutral stimuli, or conditioned stimuli (CS), like lights and bells, are through their simultaneous presentation with motivational stimuli, unconditioned stimuli (US), like footshocks or food, able to trigger a conditioned response (CR), such as freezing or salivation. The success of this learning process is measured in terms of the probability of the occurrence of a CR after the presentation of a CS. As to be expected, the reality of animal behavior in the domain of classical conditioning is more complicated than was initially anticipated (Mackintosh, 1972). In order to place the discussed models in a proper context a number of additional properties of this type of learning need to be emphasized.

At a behavioral level it seems to be appropriate to distinguish consummatory, or specific, components of learning from preparatory, or non-specific, ones (Konorski, 1967). For instance, in the case of eyelid conditioning, where a tone (CS) is presented with an airpuff to the cornea (US), the animal will display a number of responses. Next to the closing of the eye lid, which can be seen as specific to the US, non-specific behavioral or autonomic responses can be observed; startle, freezing, changes in heartrate, breathing, or Galvanic skin response. The conditioned occurrence of these non-specific responses will follow a different temporal trajectory than the specific responses. Non-specific responses show a fast acquisition (about five to 15 trials), while the development of the US specific CR takes a much larger number of trials. This behavioral distinction seems to be also reflected at the anatomical level (Lavond et al., 1993). Lesions to the amygdala, a structure in the medial temporal lobe, will strongly affect non-specific learning while lesions to the cerebellum, will selectively affect the specific learning component.

A more general interpretation of the behavior revealed in classical conditioning is that it allows behaving systems to learn about correlations between CS and US occurrences. To a certain extent, one could speak of the substitution of the US by the CS through learning. This can be seen as a crude approximation of causal relationships in the world through correlative measures (Hall, 1994; Verschure, 1996).

Operant, or instrumental, conditioning describes learning procedures in which the US is contingent on a particular action displayed by the organism. The earlier mentioned foraging experiment can be taken as an example. It was first distinguished from classical conditioning by experiments performed by Miller and Konorski in 1928 (Miller and Konorski, 1928). In these experiments a dog was trained to lift its leg in response to a cue, in order to acquire a food reward. Only when the animal displayed this required response did it receive a food reward. As opposed to classical conditioning, it is an action of the organism itself which triggers the reinforcement. The distinction between classical and operant conditioning is still debated in the field of animal learning (Mackintosh, 1972). In the work presented here we make the proposal that both phenomena reflect components which are closely coupled in the complete learning system. Both experimental paradigms address complementary subcomponents of the complete learning system.

2.2. Experimental environment

Experiments were performed using both simulated and real robots. Simulations guarantee repeatability over trials and, therefore, allow a systematic evaluation of a control structure. Only experiments with a real robot, however, allow the exploration of the robustness and generalizability of a model. The real world always being more noisy than the worst case simulation can accomplish (see Mondada and Verschure, 1993 for a further discussion and comparison of both methods).

2.2.1. BugWorld

Simulations were performed using the simulation environment BugWorld (Goldstein and Smith, 1991). In this case the simulated spherical robot (Fig. 2(a)) is using three types of sensors; range finders, collision sensors and two target sensors. The range finder consists of 37 elements distributed over 180° on the front side of the robot, each element covering a part of the range finder field. Their angular resolution decreases on the borders, 20°, and is maximal at the center, 5°. Thirty-seven collision detectors cover the same region as the range finder elements. The two target detectors are located at 90° and $-90°$ from the center of the robot. The configuration of the shape of the robot and the properties of its sensors and effectors will be referred to as the soma.

The soma can execute discrete translational and rotational actions. These atomic actions are coupled together to define behavioral patterns: ‘exploration’, ‘avoidance’ and ‘approach’. Avoidance will lead to a combination of reverse and turn actions, approach induces a turn action, while exploration induces translational motion.

Figure 2(b) displays a typical environment used in these simulation experiments. A more generalizable dimension to measure the size of an environment is provided by units of body size. In these terms, this environment measures approximately 17×10 body units. In a secluded space, multiple obstacles and targets are placed. The four targets (A, B, C, and D) each disperse a gradient which decays linearly with distance. The targets have their own dynamics.

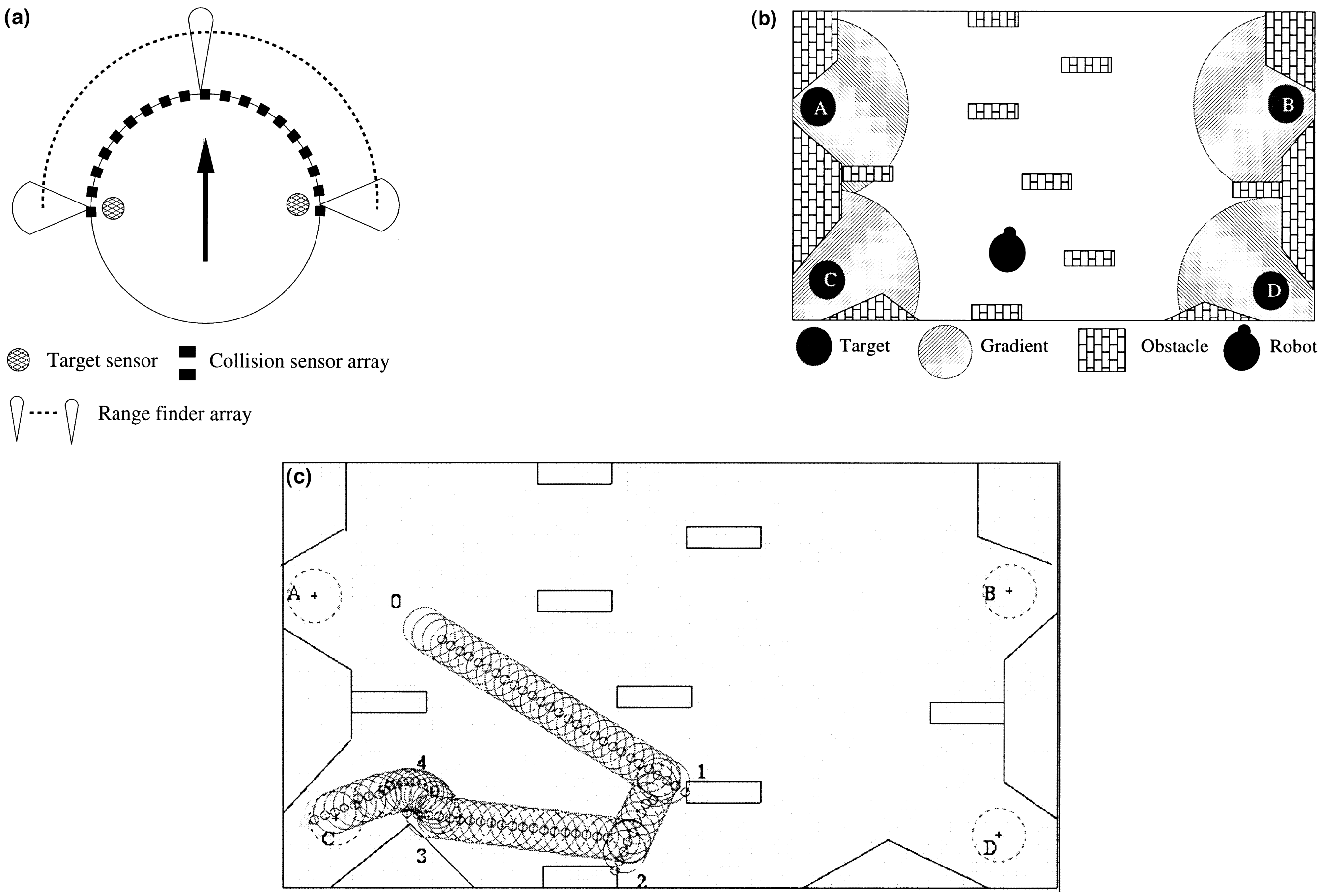

Fig. 2. BugWorld. (a) The simulated soma. (b) A standard environment containing four targets. (c) An example trajectory using a reactive control structure.

When a target is touched it is removed. A new target reappears in the same position when another target is found.

Figure 2(c) illustrates some of the behavior of the simulated robot. The positions visited are indicated with an outline of the soma. In this short trajectory a number of typical events occur. From the initial position, 0, the soma displays exploration, translational movement. Subsequently, it collides (US) three times (locations 1, 2, and 3) each time an avoidance reflex (UR) is displayed. Given the position of the collision on the soma each collision induces a turn to the right. At location 4 the gradient dispersed by target C is sensed which induces approach behaviors. The soma follows the gradient until the target is found.

2.2.2. Khepera–Xmorph

Experiments with the microrobot Khepera (K-team, Lausanne, Switzerland) were performed using the distributed simulation environment Xmorph (Verschure, 1997b).

Khepera (Fig. 3(a)) is a circular robot with a diameter of 55 mm and a height of 30 mm (Mondada et al., 1993). The basic configuration consists of two modules; the base plate and the processor module. All modules are connected by an extension bus to allow easy expansion. The base plate constitutes the elementary interface to the real world; effectors and obstacle/light detection. The robot uses two wheels for its locomotion, each wheel is driven by a DC motor. Obstacle and light detection is achieved by eight infra red send–receive sensors (IR). Six IRs are evenly placed over the front 180° of the robot and two are placed in the back. The angular resolution of the IRs is approximately 50°. The on-board computer is based on a Motorola 68331 processor with a clock speed of 16 MHz and supports 256 kByte of both RAM and ROM. Local to Khepera only the processes maintaining the serial communication, sampling of the sensors, and control of the effectors were executed. Khepera was connected to a host computer (Sun Ultra1) using a serialport at 38400 baud. Next to the two base modules Khepera was equipped with a color PAL CCD camera (K-team, Lausanne). The image from the camera was digitized with a video frame grabber (ProMovie Studio, Media Vision, Fremont, CA) attached to a Pentium-Pro PC (dual CPU 300 MHz under Linux).

The environment (Fig. 3(b)) consisted of a 90 × 60 cm secluded space (16 × 11 body units). At regular intervals along the walls red stripes were attached. In the center of the environment lines consisting of purple stripes or green rectangles were defined. A light source illuminated the

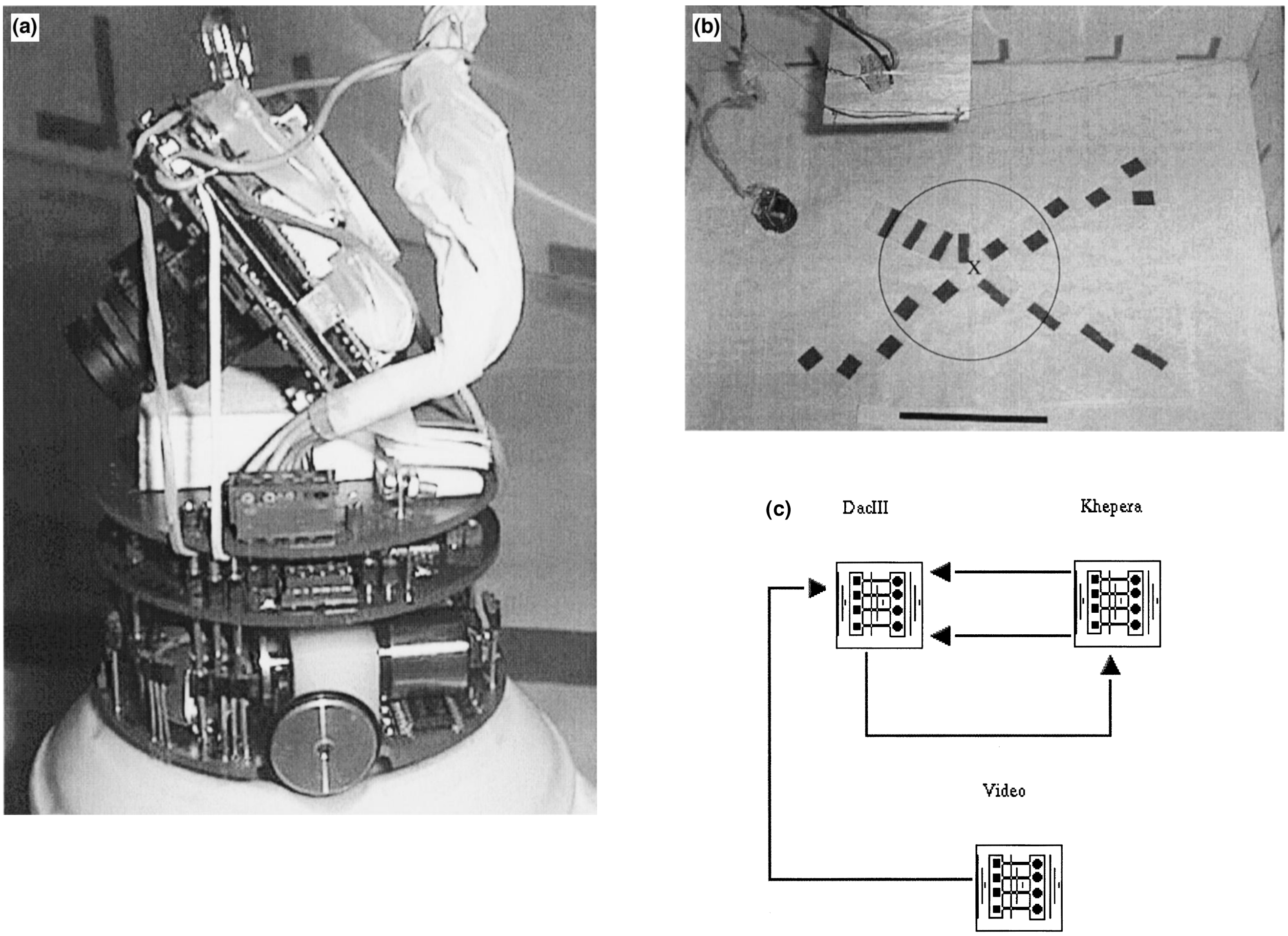

Fig. 3. Khepera and Xmorph. (a) The microrobot Khepera. (b) The used environment. Scale bar indicates 20 cm. The circle indicates the border region of the target gradient. 'X' represents the center of the target region with the highest light intensity. (c) The three simulation processes defined in Xmorph dealing with the sensors, Video, and the effectors, Khepera, and the simulation of the control structure, DacIII.

center of the lines in a region with a diameter of approximately 30 cm. Through a reflector a gradient of illumination intensity was defined.

Xmorph (Fig. 3(c)) supports the study of neural models at different levels of description. It provides a graphical specification language (using the X-Motif environment) to define, control, and analyze large scale simulations using a distributed computing method. To enhance the computational performance Xmorph uses a server-client arrangement based on the TCP/IP protocol. In this study a total of five individual, but interacting, processes were defined; front-end graphics, tracking system, and three simulation and interface processes. These processes were distributed over a LAN consisting of one Sun Ultra1 (Solaris) and four PentiumPro PCs (Linux). Processes communicated in a synchronous mode and performed at approximately 10 update cycles per second. The three simulation processes, 'Video', 'DacIII' and 'Khepera', exchange data as indicated by the connections shown in Fig. 3(c). 'Video' deals with digitizing the video image derived from the CCD camera mounted on the microrobot and the simulation of the neural system which processes the image. 'Video' exchanges the activity of a population of simulated cells reflecting the CS events (see Section 2.6) with the simulation of the control structure, 'DacIII'. In addition, 'DacIII' receives inputs from populations of simulated cells responding to US events on the robot derived from the IR sensors. 'DacIII' projects the activity of its population expressing URs to 'Khepera'. 'Khepera' in turn interprets its motor map which receives this activity and sends the appropriate commands to the robot over the serial link.

2.3. The working hypothesis on the complete learning system

Combining the assumptions on the three interacting levels of control and the distinction between the role of the non-specific and specific learning systems our sketch of a complete learning system can be further refined (Fig. 4):

1. Underlying the learning systems is an automatic system of reactive control which provides the organism with a

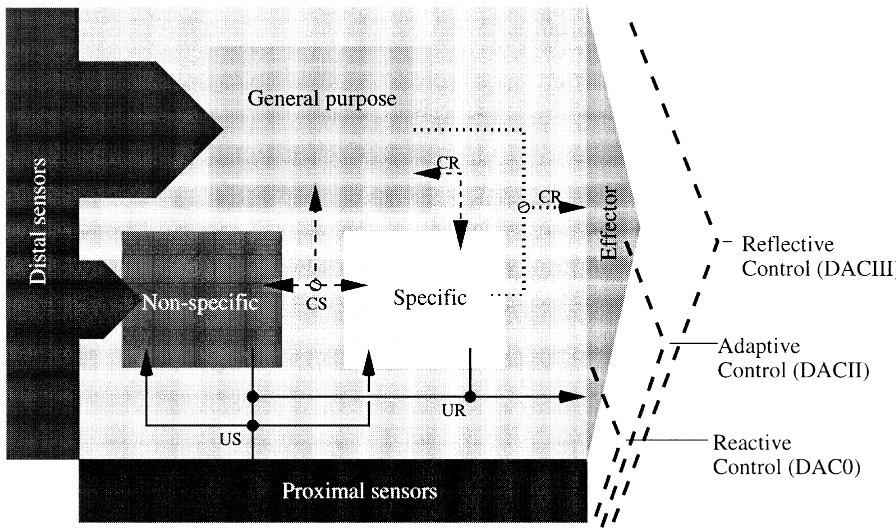

Fig. 4. The complete learning system. The assumed interactions between non-specific, specific, and general purpose components of learning and the sensors and effectors of a behaving system. Dashed lines represent operations performed on representations of the CS or CR. Dotted lines represent acquired CRs. Solid lines indicate prewired relationships.

basic level of behavioral competence. This system is fully prewired and consists of US–UR couplings. The UR can be interpreted as an expression of species specific behaviors.
2. The fast non-specific component of classical conditioning reflects the properties of a learning system which not only regulates autonomous function, preparing the organism for action, but in addition facilitates the formation of primary representations of CS events, CS identification.
3. The slow specific component of classical conditioning relates to the shaping of the CR, which is bootstrapped on top of acquired CS representations. CR shaping allows a fine tuning of predefined behavioral patterns to the actual properties of environmental challenges, i.e. timing.
4. CSs are derived from events on distal sensors (e.g. color CCD camera), while USs are derived from proximal sensors (e.g. IR sensors).
5. Operant conditioning reflects aspects of a general purpose learning system which allows the organism to form more extended representations of earlier acquired CS and CR representations, for instance reflecting their relationship in time.
6. The substrate of learning is the change in efficacies of synapses connecting different cell populations. The change of synaptic efficacy is solely dependent on the activity of pre- and postsynaptic cells, the learning process is seen as strictly local.

Components 1, 2, 3 and 4 define the adaptive control structure. The reflective control structure is defined by components 1, 2, 3, 4 and 5. In the following sections, the models of the reactive controller (called DAC0), the adaptive controller (called DACII), and of the reflective controller (called DACIII) will be described in terms of the configuration considered in the present study, in this case one CS and two US modalities. The properties of the specific learning system are not included in the present study.

2.4. Adaptive control: a model of the non-specific learning system

The control structure implementing the non-specific learning system, DACII, is based on the following assumptions (Fig. 5): (1) USs of a particular type activate specific populations of cells reflecting an internal state (IS), i.e. aversive (US^-–IS^-) and appetitive (US^+–IS^+); (2) the activation patterns in IS preserve the topology of the proximal sensor (e.g IR sensors); (3) cells in IS will activate specific reflexive actions (UR); (4) priorities between the IS

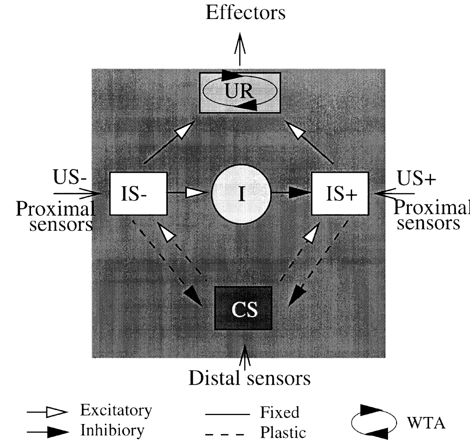

Fig. 5. Adaptive control. DACII a model of the non-specific learning system. WTA: winner take all.

populations are expressed by predefined interactions (I); (5) the CS modality (e.g video camera) is represented by a distinct population of cells preserving the topology of the sensor; (6) learning proceeds by modifying the connections between the CS and IS populations.

2.4.1. DACII: model equations describing the fast dynamics

The activity, u_j, of unit j in population CS is derived from the state, s_j, of element j of the related distal sensor:

$$u_j = j(s_j) \tag{1}$$

where j is a transduction function.

The activity of population CS is propagated to the IS populations through excitatory connections. The input, v_i^k, of cell i in IS population k is defined by:

$$v_i^k = \sum_{j=1}^{N} w_{ij}^k u_j + c_i^k \tag{2}$$

where N is the size of the CS population, w_{ij}^k is the efficacy of the connection between CS cell j and IS cell i, and c_i^k is the state of element i of US conveying sensor k. The activity, o_i^k, of cell i of IS population k is defined by:

$$o_i^k = H(v_i^k - \theta_i^k) \tag{3}$$

where H is the Heaviside or step function and θ^k defines the threshold of the units of IS population k.

The input, r_l, of unit l in the UR population is defined by:

$$r_l = \sum_{k=1}^{K} \sum_{i=k}^{M^k} y_{li}^k o_i^k \tag{4}$$

Where K denotes the number of IS populations, M^k is the size of IS population k, and y_{li}^k is the strength of the connection between cell i of IS population k and cell l of the UR population.

After updating their inputs the UR units compete in a winner take all (WTA) fashion. The winning unit's activity is again thresholded. In case its activity is suprathreshold it will induce a particular motor action. In case no motor unit is activated, the control structure will trigger exploration behavior.

A system only consisting of the US–IS and the IS–UR mapping constitutes a reactive control structure (DAC0).

2.4.2. DACII: model equations describing the slow dynamics

The learning rule employed is defined on the basis of a number of observations. In experiments with DACI (Verschure et al., 1992), a first model of an adaptive control structure, it was shown that in order to acquire and retain CS–US associations the depression component in a local learning rule needs to be activity dependent. In this way the solution reached was similar to the Oja learning rule (Oja, 1982), which is known to extract the principal components of its input set. Subsequently, it was shown that this activity dependent depression can be derived from only the postsynaptic cell, as opposed to the average activity in the postsynaptic population (Verschure et al., 1995), in order not to violate the assumption of the locality of the learning process. Verschure and Pfeifer (1992) identified two sources of instability of this local learning rule, overgeneralization and self-reinforcement. This fundamental problem was subsequently solved in DACII, without violating the assumption of the locality of the learning process, by embedding the process regulating the synaptic efficacies in a recurrent circuit. After updating the input, v^k, of the IS populations (Eq. (2)), these populations in turn recurrently inhibit the CS population. The resultant activity, u_j', of unit j in the CS population now is defined as:

$$u_j' = u_j - \gamma^{\mathrm{r}} e_j \tag{5}$$

Where γ^{r} is a gain factor modulating the effect of the recurrent inhibition and e_j is the recurrent prediction defined by:

$$e_j = \sum_{k=1}^{K} \sum_{i=1}^{M^k} w_{ij}^k v_i^k \tag{6}$$

where M^k is the size of IS population k. e will be referred to as a CS prototype.

The connections between the CS and IS populations now evolve according to:

$$\Delta w_{ij}^k = \eta^k v_i^k u_j' \tag{7}$$

where η^k defines the learning rate of the connections between population CS and IS population k.

Despite the possibility of u' to attain negative values, w is at all times kept at values greater or equal to 0. Given the effect of the recurrent inhibition this learning method is referred to as predictive Hebbian learning.

DACII will, over time, form a classification of its interaction with the environment in terms of CS events conditional to its internal states. These acquired CS prototypes, on one hand, allow the system to function as an adaptive controller and, on the other, form the representational building blocks for the construction of sequential representations. Before elaborating on the behavioral properties of DACII, the basic components of DACIII, the present approximation of the reflective controller, will be defined.

2.5. Reflective control: acquisition, retention, and use of sequential representations

The reflective controller, DACIII, inherits all properties from the reactive and the adaptive control structures, DAC0 and DACII, respectively. In addition, it contains a number of components which allow it to form and use sequential representations; the general purpose learning system.

Fig. 6 shows the central components of our present approximation of a general purpose learning system. These components deal with: (1) the acquisition of sequences of pairs of CS prototypes and related actions in a transient short term memory buffer (STM); (2) the retention of these sequences in a permanent form in long term

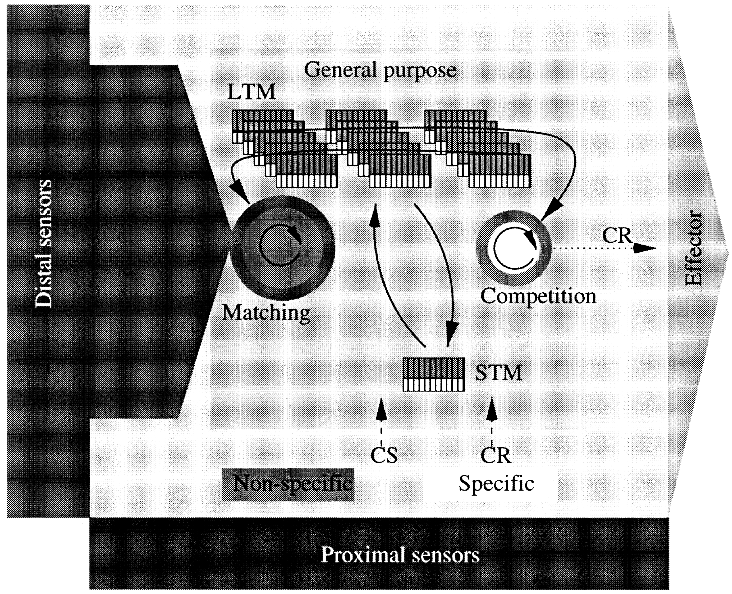

Fig. 6. The model of the general purpose learning system. Central components and their interactions are distinguished.

memory (LTM); (3) the parallel matching of retained CS prototypes with ongoing sensory events; (4) the competition between matching retained prototypes; (5) the mechanism facilitating the chaining between components of LTM; (6) the recombination of LTM components and new CS prototypes.

DACIII will bootstrap itself from a stage of reactive control to a stage of adaptive control, followed by a transition to a level of reflective control. Each transition to a higher level of control depends on constraints generated at the preceding level. In case of the transition from the reactive to the adaptive controller this constraint is provided by the actual occurrence of US events which will induce a re-mapping of the CS–IS associations (Eq. (7)). The transition from this level of control to the reflective controller depends on the quality of the matching between predicted and actual CS events expressed by an internal confidence measure, D. D depends on the accuracy of the CS prototypes formed by the non-specific learning system of the adaptive control structure. This accuracy is reflected in the result of the matching of actual, distal sensor derived (Eq. (1)), and predicted (Eq. (6)) CS events. Matching is defined by the distance, $d(u, e)$, between the feedforward generated CS activity pattern, u, and the recurrent prediction, e:

$$d(u, e) = \frac{1}{N}\sum_{j}^{N}(u_j - e_j) \tag{8}$$

D evolves according to:

$$D(T+1) = (1-\tau^D)D(T) + \tau^D d(u, e) \tag{9}$$

where τ^D defines the integration time constant.

D is a dynamic state variable which is internal to the learning system. It provides an estimate of the progression of non-specific learning and will decrease (not monotonically however) in case the constructed CS prototypes consistently match ongoing CS events. It will increase in case expected CS events are violated. This can occur, for instance, if the environment or the CS prototypes were to change for any reason.

Once D reaches a confidence threshold, DACIII will engage the general purpose learning system. In case any of the IS populations is active the generated CS prototype, e (Eq. (6)), and the related action, r (Eq. (4)), is stored in the STM buffer. This CS–UR pair will be refered to as a segment. STM functions as a ring buffer and has a finite length, N^{STM}. In case a target is found, the STM content is copied in a permanent representation, LTM. The CS prototypes stored in the LTM segments will now be matched against ongoing CS events. The result of matching is expressed in the activity of a collector unit attached to each LTM segment. The activity, $c_l(v)$, of the collector unit of LTM segment l, given IS activity v is defined as:

$$c_l(v) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{e_i}{\max(e)} - \frac{s_i}{\max(s)}\right| \tag{10}$$

where s represents the stored CS prototype. The collector units of all LTM segments interact in a competitive fashion. The probability of segment l to win this competition depends on c_l and an associated trigger unit, t_l, which acts as a dynamic threshold. The best-matching prototype minimizes the quantity $m_l(v)$:

$$m_l(v) = c_l(v)t_l \tag{11}$$

In case $m_l(v)$ of winning segment l is below a given matching threshold, its corresponding UR representation is projected onto the UR population.

Chaining through a sequence of LTM segments is defined as a probabilistic process. The activation of LTM segment 4 will increase the probability that the next segment, $l + 1$, of the sequence will be selected in the future, by reducing the value of its trigger unit $t_l + 1$; $t_l + 1 = \beta$, $0 < \beta < 1$. On each step, the activation of the trigger unit of each LTM segment decays to its default value 1: $t_l = \tau^t + (1 - \tau^t)t_l$, $0 < \tau^t < 1$.

DACIII can form recombinations of LTM segments and ongoing CS prototypes by reinserting activated LTM segments into the STM buffer.

2.6. The mapping of sensors and effectors

In case of the BugWorld simulations the cells of the CS population, $N = 37$, receive their input, s_j, from the range finder (Fig. 2(a)). The US dependent input, c^{US^+}, to the IS^+ group, $N = 2$, is defined by the sign of the difference between the states of the two target sensors. In this way, the robot can be sensitive to the gradient dispersed by a target. The IS^- population, $N = 37$, receives its input, c^{US^-}, from the collision sensors. $c_i^{\mathrm{US}^-}$ is 1 when collision sensor i is active.

In the experiments using Khepera and Xmorph both c^{US^-} and c^{US^+} were derived from the IR sensors. On average, the IR sensors will respond to reflecting surfaces placed at up to 5 cm from the sensor. c^{US^-} is defined by thresholding, θ^{CL}, the IR return signal, which gives an approximation of a collision sensor (CL). θ^{CL} was set such that only surfaces closer than 1 cm from the sensor would trigger suprathreshold activity. The raw IR signal was projected onto a population of leaky integrator linear threshold units, $N = 6$, which rendered c^{US^-}. c^{US^+} was derived from the ambient light (AL) detected by the IR sensors in their passive mode. This signal was projected onto a population of leaky integrator linear threshold units, $N = 6$. Their activity was thresholded, θ^{AL}, in order to reduce the background, level of ambient light. By thresholding, θ^{T}, AL with an appropriate value a measure is defined which reflects the presence of a target.

The dynamics of both US populations are defined in similar terms. The membrane potential, vm_i^{US}, of US transducing unit i is defined as:

$$vm_i^{\mathrm{US}} = \beta^{\mathrm{US}} vm_i^{\mathrm{US}} + \gamma^{\mathrm{IR}} IR_i \tag{12}$$

where β^{US} specifies the decay rate of vm^{US}, γ^{IR} the excitatory gain due to the IR signal, and IR_i the return signal of IR_i, either in active or passive mode.

The activity, c_i^{US}, is defined through thresholding the integrated input:

$$c_i^{\mathrm{US}} = H(vm_i^{\mathrm{US}} - \theta^{\mathrm{US}}) vm_i^{\mathrm{US}} \tag{13}$$

The multiplication of the Heaviside with vm_i^{US} is only applied to AL. Motor output sent to Khepera is derived from a topologically structured map as used in earlier work (Verschure et al., 1995). Continuous rotational or translational motion is defined by patterns of activity in population M which consists of leaky integrators, $N = 100$. The units in M receive external excitatory inputs from the UR population. The pattern of innervation is specific for each UR unit, since they each represent a specific behavior. The units in M update their membrane potentials following Eq. (12) to which now an inhibitory input is added derived from all other units in M. When the winning unit is above threshold, θ^M, it will define the motor commands the robot will execute. Note that as opposed to the simulation in this case motor activity is continuous, once initialized the motors will only change their state in case another pattern of activity arises in M.

The distal sensor, which defines CS events, was provided by the color CCD camera mounted on Khepera. The 480 × 640 image was compressed to an image size of 210 × 210. Every color channel of the digitized image, using a RGB representation, was pixelized (reduction ratio: 4 × 4:1) onto a distinct population of leaky integrators, $N = 2500$, conserving the ‘retinotopy’ of the camera. Their membrane potentials and activity were updated according to Eqs. (12) and (13). The population conveying the CS states, $N = 36$, was subdivided into three subregions, each cell reflecting the relative dominance of a particular color channel in particular regions of the image. Each unit received excitatory input from a topologically mapped (15 × l5) region of the preferred color channel and inhibition over a wider surround (30 × 30) in the two opposing color channels. The membrane potential, vm_i^C, of cell i of population C is defined as:

$$vm_i^C = \beta^C vm_i^C + \gamma^p \sum_p^{N^p} w_{ij}^p c_j^p - \gamma^{o1} \sum_j^{no1} w_{ij}^{o1} c_j^{o1} - \gamma^{o2} \sum_j^{no2} w_{ij}^{o2} c_j^{o2} \tag{14}$$

where β^C specifies the decay rate of vm^C, γ^p the gain of the preferred color channel, c_j^p the value of pixel j of the preferred color channel p, and w_{ij}^p the connection strength of the projection of cell j to cell i. Indices o1 and o2 refer to the two opposing color channels.

The activity, s_i^C, of unit i is defined through thresholding the integrated input:

$$s_i^C = H(vm_i^C - \theta^C) vm_i^C \tag{15}$$

Fig. 7 shows the properties of the model processing the color image and producing the mapping to the CS population. Fig. 7(a) shows the projections onto one representative cell of each color region in the CS population. Fig. 7(b) displays the configuration used to illustrate the response properties of the CS modality in which a red rectangle was placed in front of the robot. Fig. 7(c) represents the compressed image derived from the camera using a standard luminance to gray mapping. Fig. 7(d) shows the response of the three color channels to the stimulus and the response of the CS population. In this case a single cell in the region of the CS population responding to red is active. Through balancing the excitation from the prefered color channel with the inhibition from the two opponent channels a

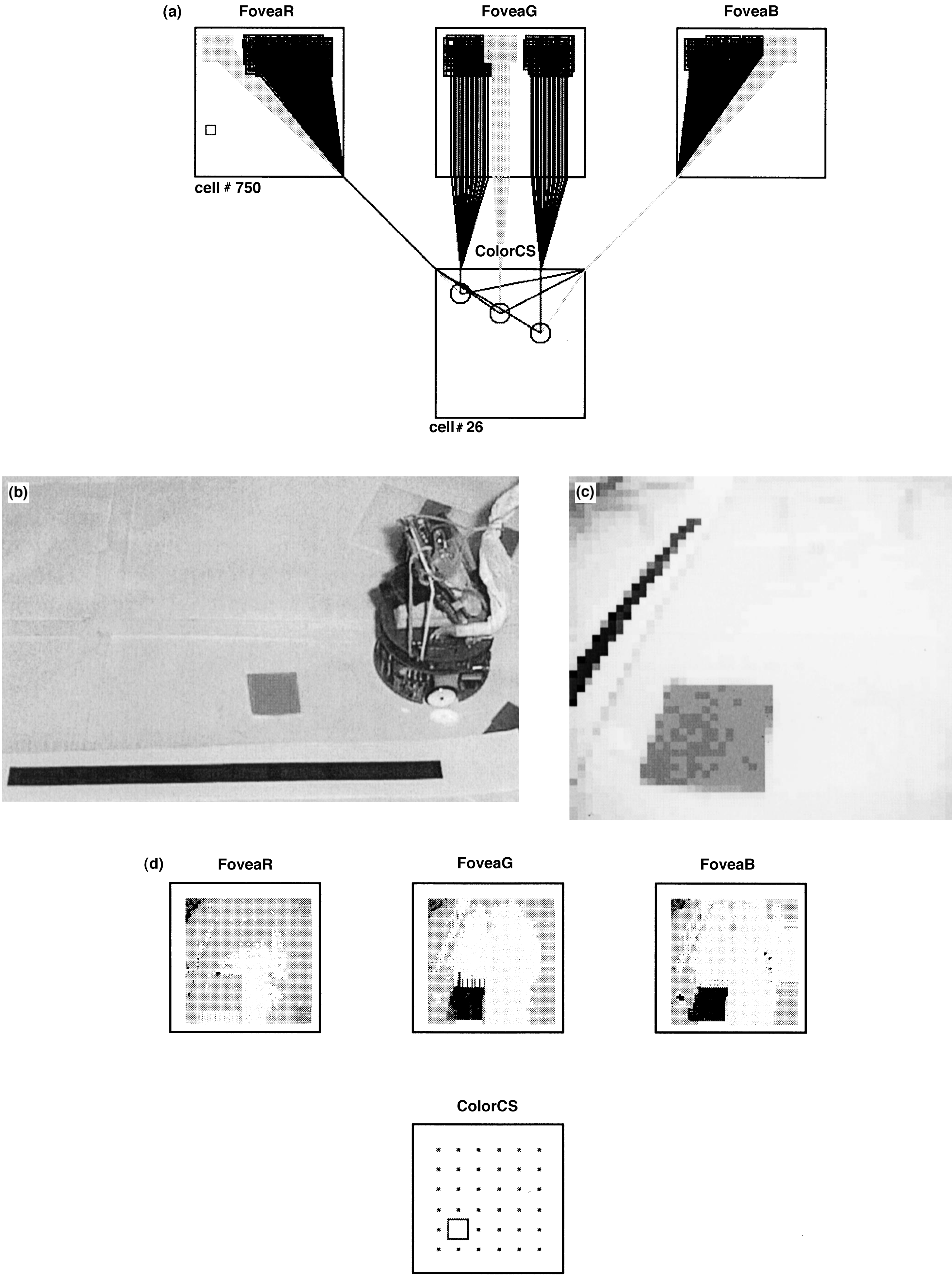
(a)
FoveaR
FoveaG
FoveaB
cell # 750
ColorCS
cell # 26
(b)
(c)
(d)
FoveaR
FoveaG
FoveaB
ColorCS

Table 1
A performance comparison of DAC0, DACII, and DACIII

Control	Targets	Collisions	Traveled distance	Collisions/ targets
DAC0	53	532	66 170	10.04
DACII	34	106	60 590	3.12
DACIII	53	73	39 910	1.38

robust response to colors can be achieved over a range of illumination conditions.

3. Results

By means of the simulated robot the basic properties of both DACII and DACIII will be illustrated. The experiments with Khepera serve to demonstrate that the proposed model generalizes in a straight forward manner to a real robot. Before turning to a more detailed analysis of DACII and DACIII, a performance comparison of the three forms of control distinguished will be described.

3.1. A comparison of the three models of control

In order to delineate the performance difference between the three types of control, reactive (DAC0), adaptive (DACII), and reflective (DACIII), all three models were applied to the same task of finding targets in an environment containing multiple obstacles. In this simulation experiment the environment depicted in Fig. 2(b) was used. The robot could explore this environment for a total of 7000 time steps. The target gradients were only present for the first 2000 time steps. In this way, a recall period, lasting 5000 time steps, was introduced. In this period, the robot either finds a target through the use of acquired representations or by coincidence. Table 1 summarizes the performance of the three forms of control.

Table 1 shows that there is a strong performance difference between the three forms of control. DAC0 finds a significant number of targets, but suffers a high number of collisions. The overall collision to target ratio is 10.04 and the traveled distance is 66 170. DACII reduces the number of collisions compared with DAC0, but finds less targets. DACIII further reduces the number of collisions and finds as many targets as DAC0. In addition, its total traveled distance is markedly lower than for the other two control structures. To further exemplify the performance difference between DACII and DACIII. Figure 8 displays the trajectories of both control structures during the recall period.

In the recall period, DACII does not collide with any obstacles anymore, as a result of previous learning experiences. The displayed trajectory, however, shows that its behavior is highly variable. DACII practically covers all positions in the environment. Since its actions are reactive to immediate sensory events, CS or US, little temporal structuring of its behavior can be observed. This is in sharp contrast to DACIII which has settled into a trajectory which is highly regular and approximates the shortest route between the different targets. This suggests that it has created sequential representations which seem appropriate for the present task. The structuring of the behavior, through the use of the general purpose learning system, also explains the reduced number of collisions DACIII suffers as opposed to the other control structures. Since DACIII covers a reduced region of the environment the probability to encounter obstacles also decreases. The relatively low value of the traveled distance of DACIII can be explained in terms of the properties of the behavioral stereotypes; avoidance, and approach. Approach behaviors have no translational component, hence the more a control structure is, directly or indirectly, under the influence of population IS^+ the less its traveled distance will become. This indicates that DACIII to a large extent relies on sequences containing approach behaviors.

3.2. The dynamics of the confidence measure D

The transition from adaptive control to reflective control depends on the internal confidence measure *D* (Eq. (9)). The performance test described above demonstrated that DACIII did reach its confidence threshold and engaged the general purpose learning system. Figure 9 provides a more detailed description of the dynamics of *D*.

The performance of DACIII in this experiment was equivalent as in the earlier described performance comparison. Figure 9 shows that *D* rapidly decreases over the first 2000 time steps. At the onset of the first recall period *D* transiently rebounds and subsequently shows a practically constant decrease with time. When the target gradients return at time step 7000 this decrease is accelerated. *D* reaches an asymptotic level after approximately 8000 time steps.

Together with the performance of DACIII (see Fig. 8(b)), this implies that the internal confidence measure *D* reliably reflects the quality of acquired CS prototypes. *D* shows

Fig. 7. Properties of the modeled sensory system processing states of the distal sensor (color CCD camera). (a) An illustration of the projections between the three populations responding to the color channels and the CS generating population. Light gray lines indicate excitatory connections, dark gray is inhibitory. (b) Khepera placed in front of a red rectangle. Scale bar is 20 cm. (c) The digitized video image using a standard hue to luminance mapping. (d) A single cell in the CS population responds to the red rectangle present in the image. Only for this cell the excitation, derived from the preferred red channel (population FoveaR), exceeds the inhibition received from the two opposing color channels, green and blue (populations FoveaG and FoveaB). Levels of activity are expressed in a gray scale and the size of the rectangles representing the individual cells. Light gray and large rectangles represent maximum activity, dark gray and dots represent minimum levels of activity.

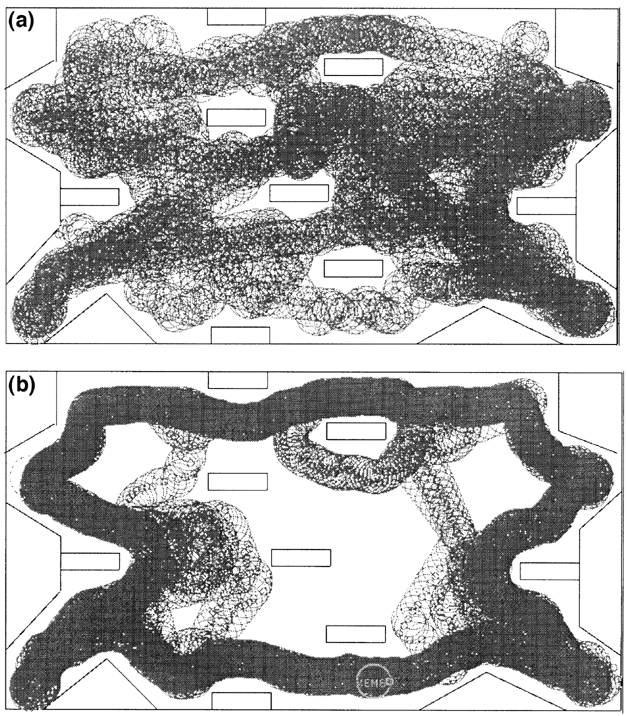

Fig. 8. Performance comparison in the recall period. (a) Trajectory of DACII. (b) Trajectory of DACIII performing the same task.

that the matching between the ongoing events on the distal sensors progressively improves. In addition, Fig. 9 suggests that it can be considered as an implicit time indicator.

3.3. The acquisition and use of sequential representations

Figure 8(b) showed that DACIII is able to display a highly structured behavioral trajectory over extended periods of time. The underlying LTM segments, however, do not necessarily need to directly reflect this coherence. This raises the question of the content and relationships of the sequential representations that affected the performance. As a first approximation of the analysis of the LTM segments we can pose the question in what position in the environment effective LTM segments, that matched ongoing CS events and induced actions, were actually stored in the STM buffer. The distribution of these locations provides a measure of the specificity and the coherence of the LTM representations.

Figure 10 displays this acquisition distribution for the experiment with DACIII described in Fig. 8(b). Every time a LTM segment induced an action, the position were it was stored in STM was plotted with the outline of the soma. The spatial distribution of the acquisition of effective segments shows that most were acquired in four specific regions in the vicinity of the four targets. At each target distinct approach sequences were acquired which captured the detailed differences at these four locations. These frequently reused sequences, which are most densely labeled, fall mostly within the region of the target gradient. A second type of effective segments, however, were acquired outside of these gradient regions. These are of particular interest. These segments were acquired when learned approach or avoidance actions were displayed. This demonstrates that not only the content of the CS prototypes depends on the learning experience, but that also their inclusion in LTM segments reflects the learning history. Comparing with the actual trajectory displayed by DACIII (Fig. 8(b)) shows that this latter type of sequences were generalized to other situations. This analysis shows that DACIII has parcelated its representation of its interaction with the environment in terms of a limited and coherent set of prototypical situations.

This provides a possible scenario for explaining aspects of the foraging behavior. Figure 8(b) showed that DACIII followed a linear strategy. The interpretation of the used LTM representations indicated that this linear strategy is based on a limited set a prototypical situations defined in terms of the motivational state, appetitive, and the CS prototypes and their associated actions. Hence, a continuous representation of the complete trajectory is not required to induce this highly structured behavior. In addition, globally structured behavior can be achieved through the use of local, egocentric views of the world. This property of DACIII can be partly explained through the generalization of particular

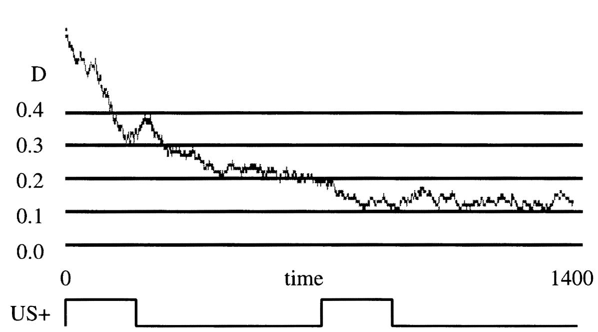

Fig. 9. The confidence measure D. Evolution of D of DACIII over 14 000 steps using the environment depicted in Fig. 2b. The target gradients were present from time steps 0–2000 and 7000–9000 (see lower panel).

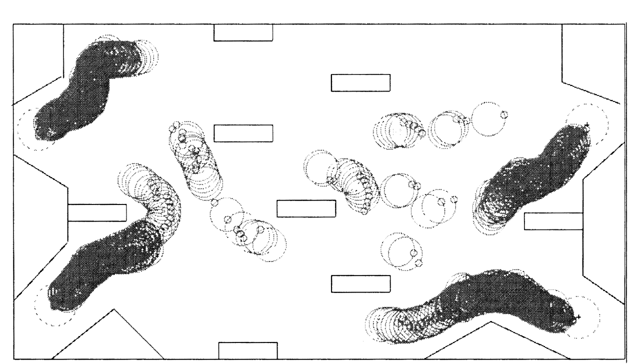

Fig. 10. Positions in the environment where effective LTM segments were stored in STM.

sequences to other positions in the environment, but especially by the emphasis of the mechanisms for acquisition and expression on events which deviate from default behavior. This aspect of DACIII's behavior suggests, therefore, that not only in the interpretation of specific sensory events generalization can be achieved, but also in the formation and especially expression of more abstract sequential representations, which combine both sensory and effector components.

3.4. Results with Khepera–Xmorph

In the experiments with the microrobot Khepera the aim was to demonstrate that DACIII generalizes to a real-world device using sensors and effectors with very different, and certainly less ideal, properties than the simulated device. In these experiments the environment depicted in Fig. 3 was used. The position of the robot was tracked using a ceiling mounted PAL CCD camera and the tracking module, TraX,

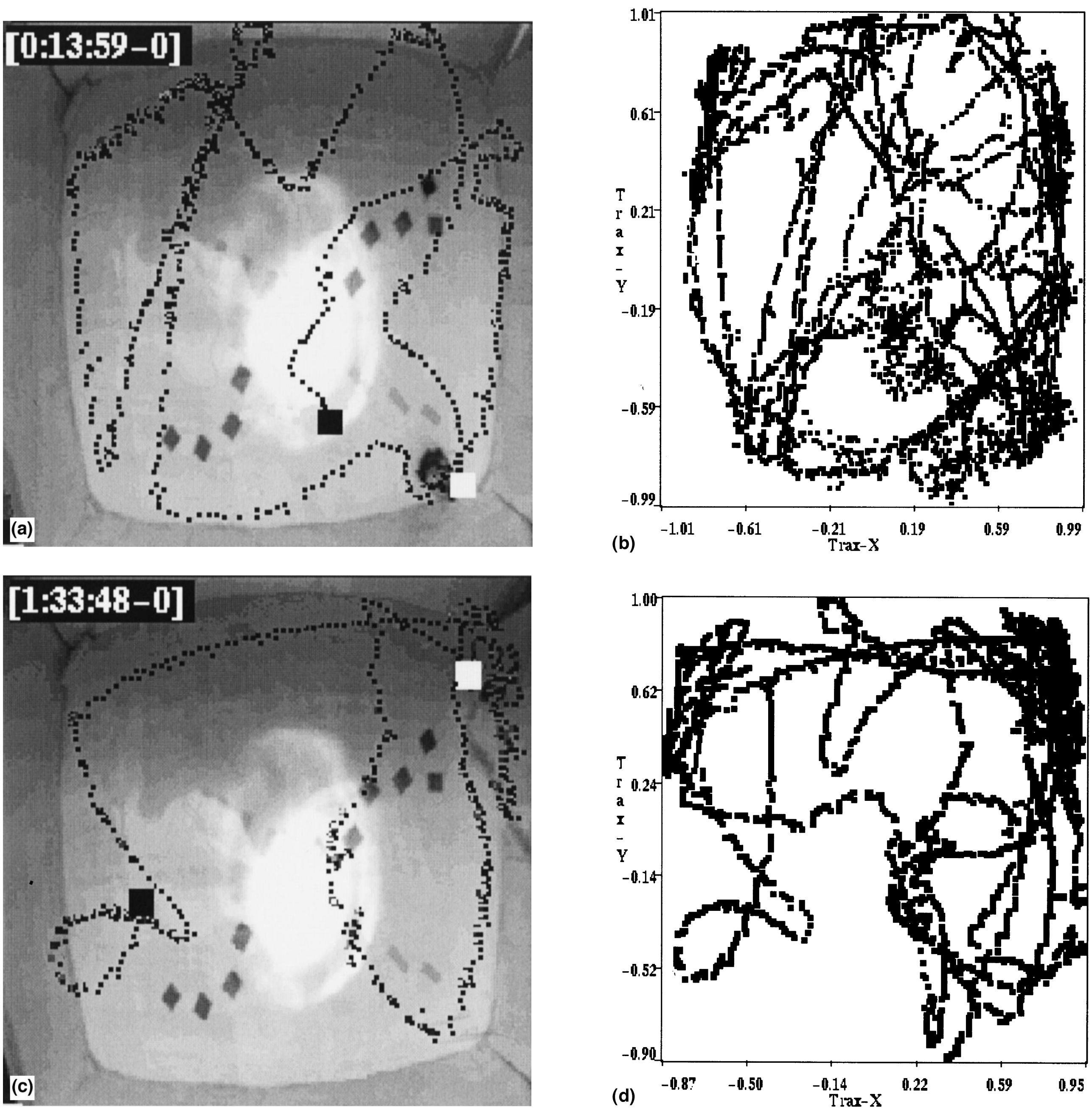

Fig. 11. Performance of Khepera using DACIII. Time intervals are defined as hours:minutes:seconds. (a) Example trajectory in time interval 00:10:45 and 00:13:59. Individual points in the plot reflect the position of Khepera as sampled through TraX. The white and black rectangles represent the position of the soma at the start and end of this sequence respectively. (b) Positions visited by the soma during the first 26 min of the experiment. (c) Time interval 1:29:57–1:33:48. (d) Positions visited by the soma in the time interval 1:08:55–1:33:48.

of Xmorph. In addition, relevant state variables were continuously logged. Fig. 11 displays the performance of the model in a trial that lasted a total of 2 h. The model used its first LTM segment after 24 800 cycles which is equivalent to 48 min.

Figure 11(a) shows a typical trajectory in the early stages of learning. This trajectory was generated during 3 min and 14 s beginning at 10 min and 45 s after the start of the trial. Khepera started out at the lower right corner of the environment, indicated with the white rectangle, and finished approximately 3 min later at the lower side of the target region, black rectangle. In the early stages of learning, the behavior is dominated by reactive control and progressively by adaptive control. In this period the behavioral trajectory, summarized for the first 26 min in Fig. 11(b), is characterized by periods of translational movement deflected by collisions, and avoidance actions, or the detection of the target gradient, accompanied by approach actions. After the transition to reflective control (Fig. 11(c) and (d)) the translational motion is regularly interrupted by sequences of actions induced by the reflective control structure. The LTM representations, in turn, are activated by the colored markers on the floor and the walls of the environment. This is illustrated in detail for the trajectory displayed in Fig. 11(c) in Fig. 12.

Figure 12 displays the positions visited by Khepera in a time interval starting at 89 min after the start of the trial and which lasted about one minute. The positions visited by Khepera where actions were defined by the reflective control structure are indicated with rectangles. Positions in the environment where a target was found are indicated with stars. Early in this trajectory, in the vicinity of the green rectangles attached to the floor of the environment, several subsequent actions are under reflective control. The perceived green rectangles matched with some of the CS representations stored in the LTM segments. Subsequently, the robot moves towards the wall and collides. After turning into the open field another collision occurs with the upper wall. While crossing the set of green rectangles reflective control is activated and the green rectangles are followed for a number of steps. A few seconds later this reoccurs. In this case, reflective control remains active for 13 consecutive time steps and induces a turn towards the target region. Subsequently, the target is found. This sequence of actions demonstrates that the non-specific learning system has constructed stable representations of CS events which the reflective control structure has combined, with their accompanying actions, in an appropriate way in LTM.

As a second example of the ability of DACIII to successfully control a real-world device, a set of experiments were

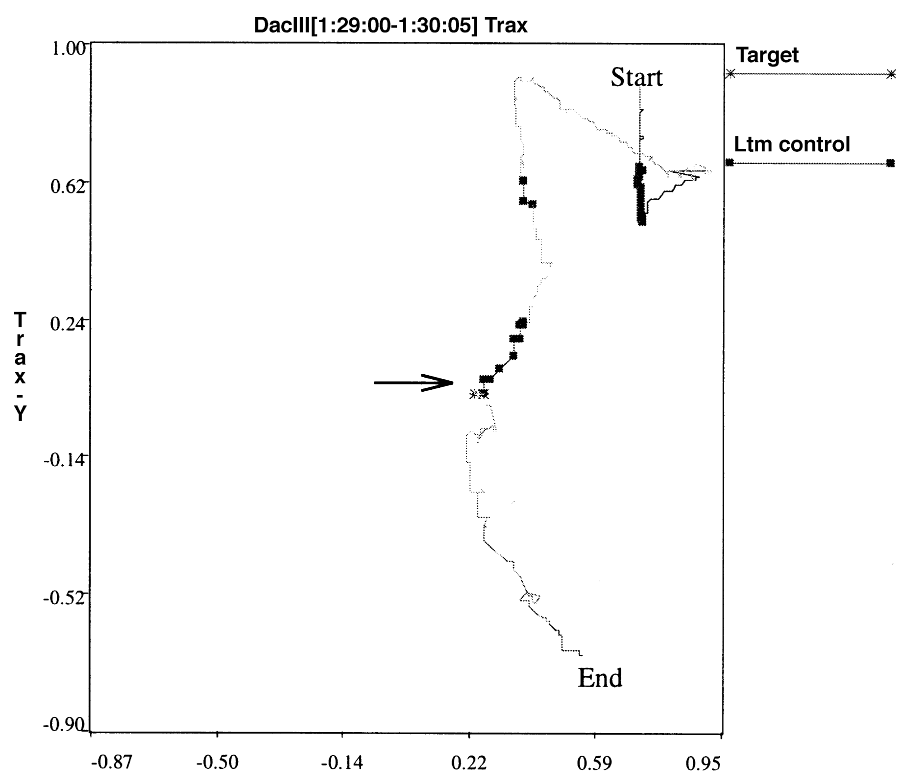

Fig. 12. Illustration of the structuring of the behavior of Khepera through the use of sequential representations. Positions where the behavior was determined by reflective control are indicated with a rectangle. The location of the robot where it found a target is indicated with a star. The start and end position of the soma in this interval is indicated with 'Start' and 'End'. The arrow indicates a situation where under the continuous control of the internally generated predictions a target was found.

performed using a similar environment (Fig. 13(a)). This environment measured 37 × 57 cm. Next to red stripes attached to the wall a red triangle was placed on the left center part of the floor to evaluate the avoidance responses acquired by the adaptive controller. A path of green or purple rectangles was leading to the target region. The trajectory of the first 45 min (Fig. 13(a)) demonstrates that the red triangle is systematically avoided and that the target region is regularly visited. In the recall test the light source was switched off and the robot was repetitively placed in the upper left corner of the environment, marked with a white rectangle. The orientation of the robot was such that it would not be able to reach the target region through translational motion only. In all trials the target region was visited (Fig. 13(b)).

Both through a direct analysis of the relationship between the performance of DACIII and the effectiveness of reflective control and a recall test it is demonstrated that the presented model of a complete learning system generalizes well to a real-world device.

4. Discussion

The aim of this paper was to describe a model of a complete learning system which could provide a heuristic in understanding the forms of behavior displayed in, for instance, a foraging task. The presented approximation of a complete learning system demonstrated that aspects of these forms of behavior can be understood in strictly bottom up terms. Using reactive control as a foundation for learning the experiments described showed that an adaptive control structure can be defined which extracts representations of CS events out of the interaction between the soma and the environment. The representations of CS events, called CS prototypes, express a relationship between a particular state of a distal sensor and an internal state. Through this coupling of a sensory event and an internal state, implemented by the synaptic efficacies of the projections between the CS and IS populations, the CS representation is implicitly associated with a particular behavior. Hence, three components of a CS representations can be distinguished: its content derived from the state on the distal sensor (the CS properly); its meaning defined by the internal state (in the present case appetitive or aversive derived from the encountered USs); and an action pattern (UR). The presented model of adaptive control suggests that the construct representation needs to be considered in terms of these three closely coupled components. In addition, this model demonstrated that the process of CS identification can be based on a fully local learning rule. Subsequently, reflective control, using sequential representations of CS prototypes and UR representations, can be bootstrapped on top of the adaptive control structure. The reflective control structure, in turn, is able to induce highly structured forms of behavior. This structuring, owing to the chaining mechanism, in turn is

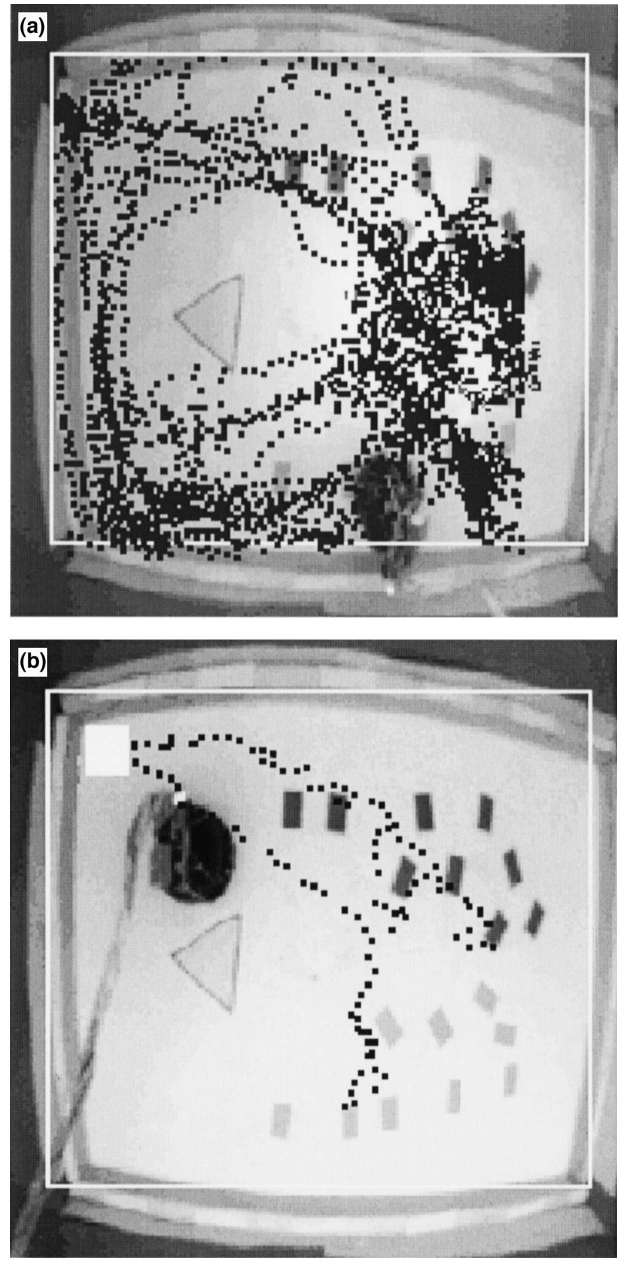

Fig. 13. Illustration of the structuring of the behavior of Khepera through the use of sequential representations in a recall test in a different environment. (a) Positions visited during the first 45 min. (b) Positions visited during three test trials where the robot was placed in the upper left corner of the environment indicated with the white rectangle.

only defined in terms of the local interactions between the segments which form the sequential representations. Activated segments affect future classifications only by transiently increasing the probability that the subsequent segment in the sequence will dominate the competition process implemented by the collector and trigger units. This allows the reflective control structure to dynamically construct and maintain multiple ‘plans’ for its behavior. It is this property of our present model which could be interpreted as reflective.

In developing this model, the assumption was made that complementary components of the complete learning system are revealed through the paradigms of classical and operant conditioning. Our results demonstrate that this is a feasible option. The processes revealed through classical conditioning, adaptive control, laying the foundation for the processes studied through the paradigm of operant conditioning, reflective control. Before elaborating on the implications of the results some elements of the presented models will be discussed.

DacIII is presented as a first approximation of a complete learning system and a reflective control structure. At this stage of its development, however, it is not claimed that it actually is complete. Many elements are still missing, and provided our only limited understanding of the behavioral and the neuroscientific domains some elements still await their specification. DACIII does provide a first step towards the study of these systems and allows a systematic exploration of scenarios dealing with the domain of operant and classical conditioning which facilitate an interaction between these two domains of inquiry.

In this study, we choose a formulation of the predictive Hebbian learning mechanism which can induce negative levels of activity in the CS population as opposed to the original definition (Verschure and Pfeifer, 1992). This choice was based on the wish to find a smooth approximation of the asymptotic values of the connection strengths. This precludes a direct application of this method as a heuristic in the study of biological systems. The method of predictive Hebbian learning, however, does require further study. It, for instance, replicates the observed response properties of the ventral tegmental area (VTA) (Schultz et al., 1997). It has been shown that the dopaminergic cells in this region show an enhanced response, to background, in anticipation of rewarding events, which in turn can be suppressed below background in case the anticipated reward does not occur. In addition, an equivalent method has been successfully applied to the study of cortical dynamics (Rao and Ballard, 1997). In current work we are exploring the option to allow the recurrent inhibition of the CS population to change the level of activity given a particular level of background activity. This implies, however, that the dynamics of the weights needs to be extended with a variable threshold as proposed in (Bienenstock et al., 1982). In the case of predictive Hebbian learning, however, the dynamic threshold will express the presynaptic drive onto a particular synapse as opposed to the time averaged post synaptic activity. Preliminary results have shown that this is a feasible option.

The main problem which has not been explicitly addressed in the present study is how STM and LTM representations are retained. The current version of DACIII relies on algorithmic solutions. The distinction between specific brain structures involved in either acquisition, amygdala, or retention, cortex, needs to be made in the study of learning and memory and in the proposed model. Yet, no clear proposals are available how this transformation is accomplished. This is an open problem which will take a central role in the further study of a complete learning system.

Many models have been proposed dealing with either classical or operant conditioning (i.e. Klopf, 1982; Sutton and Barto, 1981; Grossberg and Levine, 1987; Grossberg and Schmajuk, 1987). As opposed to these models the DAC modeling series, which has its background in a model of classical conditioning Verschure and Coolen (1991), took as its central theme the problem of the acquisition of CS representations, or CS identification, which was proposed to be one of the central elements of the learning system studied through the paradigm of classical conditioning. These alternative approaches, however, were focused on the acquisition of CS–US or CS–UR associations assuming that the respective CS, US and UR representations are given a priori. DAC also deviates from the main stream of models studied in the domain of machine learning (see Kaelbling et al., 1996 for a review) by its insistence on local learning methods. DAC confirms, however, Grossberg’s hypothesis (Grossberg, 1982) on the importance of distinguishing the effects of a short term drive representation, in DAC terminology the internal state, and the CS representation in the explanation of classical conditioning. This distinction, however, has roots both in the study of behavior (Konorski, 1967) and the neuroscience of learning (Thompson et al., 1983). Armony et al. (1995) proposed a model of classical conditioning which described the development of receptive field properties of thalamic and cortical cells induced by fear conditioning. This model, which relates to the properties of the adaptive controller (DACII), only provides a very abstract description of these dynamics. It does provide an additional example, however, of the hypothesis put forward by the DAC series that the observed effects of classical conditioning on the autonomous nervous system only provide a restricted picture on the role of the non-specific learning system. Traditionally, the role of classical conditioning has been defined in terms of the acquisition of CS–US associations. Its effects should be expanded, however, to include the dynamic formation of CS representations. This is also suggested by the physiology of both the primary auditory (Weinberger et al., 1993) and visual (Galuske et al., 1997) cortex in conditioning tasks. It has been demonstrated that neurons in both areas, conveying distal sensor information, can adapt their tuning curves to reflect the properties of a CS.

Based on the method of convergent validation, the subsequent models in the DAC series have been extensively studied using both simulated and real robots and a wide range of sensor and effector systems (Verschure et al., 1992, 1995; Verschure and Pfeifer, 1992; Almassy and Verschure, 1992; Mondada and Verschure, 1993). This aspect of DAC can be best compared with the work on the mobile robot MAVIN (Baloch and Waxman, 1991). Despite its relatively restricted focus on visual object recognition it is one of the first examples of a complete control structure applied to a mobile robot based on observations derived from the behavioral literature. A model of operant conditioning, applied to a delayed match to sample task, implemented on a robot has been proposed (Touretzky and Saksida, 1996). This model, as opposed to DACIII, is aimed at a functional decomposition of the task at hand, using a production system implementation, and does not allow any cross validation with a neuroscientific level of description. As such it faces the problem of indeterminancy pointed out in the introduction and its application to a real-world device does not seem a necessary component in understanding the proposed functional decomposition.

Several models dealing with sequence learning have been proposed. On one hand a large number of these models are derived from Hopfield networks (Hopfield, 1982) which include a transduction delay (e.g. Morita, 1996). In our earlier work on classical conditioning we have demonstrated that these types of networks can be successfully applied to the modeling of both delay and trace conditioning (Verschure and Coolen, 1991). In the case of the acquisition, retention, and expression of sequential representations, however, these models are not sufficient. DACIII shows that an important component of the complete learning systems is the parallel matching and competition of LTM segments and the expectancy dynamics implemented by the collector and trigger units. In order to implement such a system, CS prototypes need to be represented as distinct entities in the underlying substrate. Hopfield networks, however, would represent the CS prototypes as attractors which cannot be guaranteed to be distinguishable at any one point in time. Hence, they do not provide a feasible option. A second class of models explicitly addresses the biological substrate involved in sequence learning (e.g. Dominey et al., 1995; Denham and McCabe, 1995; Dehaene and Changeux, 1997). All these models emphasize the close interaction between frontal cortex and the basal ganglia and imply a system implemented by the STM-LTM dynamics of DACIII. In all cases, however, the CS identification problem has been side stepped and the models have not been evaluated in terms of behaving systems. This can account for the different solutions pursued. For instance, DACIII relies strongly on the internal confidence measure *D*. It was argued that such a variable expressing the ability of the learning system to reliably classify its interaction with the environment is a necessary component of a complete learning system. It can be seen as a gating signal for the acquisition of STM representations. The proposed confidence measure, does provide an hypothesis on the type of state variables that a reflective control structure, such as a mammalian brain, needs to maintain in order to function effectively. Both alternative proposals mentioned lack such a measure. They also lack a clear framework specifying how CS representations are acquired and retained. As DACIII both proposals, however, emphasize the importance of the continuous matching and competition between representations. In this case, the matching is interpreted as a process implemented in frontal areas of the neocortex, while the competition is implemented through the cortico-basal ganglia loop. In the further development of the DAC series, the different components of the proposed model are replaced with models which reflect more closely the anatomical and physiological properties of these brain areas (e.g. Verschure and König, 1997). Only after this modeling exercise can we with more confidence provide anatomical labels to the subcomponents of DACIII, i.e. functional components distinguished in a model do not necessarily map directly and uniquely onto specific brain areas. At our present level of modeling it seems more appropriate to not violate the obvious, i.e. by insisting on local learning methods, as opposed to too quickly generalize the putative models to highly intricate and still only partly understood brain structures.

In the present version of DACIII the complexity of the CS representations are severely reduced compared with what biological systems can accomplish. This implies that the actual behavioral implications of the models cannot be fully explored. The issue of learning is closely tied to the notion of representation. In addition, as mentioned earlier, the model components are defined in too abstract terms to allow a validation against neuroscientific data which the method of convergent validation prescribes. In order to alleviate this situation a parallel modeling effort dealing with the way in which cortical circuits can form dynamic, spatial and temporal scale, invariant representations has been performed which includes pertinent anatomical and physiological features of cortical circuits (König and Verschure, 1995; Verschure and König, 1997). In addition, in order to arrive at more biologically realistic real-world devices, initial experiments were performed using neuromorphic sensors (silicon retinae) as distal sensors on mobile platforms (Indeveri and Verschure, 1997). These sensors approximate the response properties of the outer plexiform layer of the retina (Douglas et al., 1995). They provide an input signal which emphasizes the dynamics of the visual world, rapidly adjusting to changing illumination conditions and responding to spatio-temporal contrast variations. Hence, these distal sensors provide more realistic constraints on neural models which are supposed to work with these signals as opposed to CCD cameras.

An important question is whether the proposed model, which captures elements of problem solving tasks such as foraging, can be considered a model of cognitive processes. The dominant paradigm in the study of mind, brain, and behavior can be called symbolic cognitive psychology (Newell, 1990). This approach bases its explanations of cognition on a so-called knowledge level. A central principle in a knowledge level explanation is the law of rationality: a rational system will use its knowledge in order to reach its goals. A paradigmatic example of this approach, which constituted the core of the artificial intelligence program, is the hypothesis of physical symbol systems (PSS) put forward by Newell and Simon (Newell, 1980). Despite its limitations the proposed model of the reflective controller, DACIII, is the closest approximation of a synthetic rational system, which uses its knowledge to reach its goals. The goals are defined in terms of its internal states, i.e. avoid or approach. When the IS population aversive is active, for instance, the adaptive control structure will aim the behavior of the system to the reduction of this internal state, i.e. by triggering avoidance actions. As such, both the avoidance of obstacles and the approach of targets can be interpreted as goals the system tries to attain. The reflective control structure is, in addition, attempting to achieve the goal of finding targets. The knowledge it brings to bear on reaching these goals are the acquired LTM segments, which can be interpreted as the world model of the system. This world model, however, is at no point in time fixed. The content of LTM can change at any time owing to new experiences (see Verschure (1998) for a further comparison). Traditionally the ascription of a goal to a behaving system is defined in terms of performance. The presented model of the reflective control structure makes the proposal that its neuronal correlate will have a component which relates to the motivational state of the organism. As such, the definition of a representation in terms of a sensory event, an internal state, and an action implies that the notion of a goal is an integral component of the acquired CS representations.

DACIII is a fully bootstrapped system. Initially, it performs as a reactive controller which provides the constraints to develop CS representations. Through the acquisition of these CS representations the system will start to behave as an adaptive controller. Subsequently, the transition to reflective control can be made in case the non-specific learning system reliably classifies the ongoing interaction between the organism and the environment. At this level the developed CS prototypes can start to function as expectations on future states of the world expressing their relative confidence in terms of the dynamics of the collector and trigger units. These expectations will, in turn, strongly structure the actual behavior displayed. Even though many problems remain to be solved, DACIII demonstrates that also more complicated, ‘cognitive’, problem solving tasks are within reach of a pure bottom up approach, the reservations of the cognitivists not withstanding (Fodor, 1983).

References

Almassy, N. & Verschure, P. F. M. J. (1992). Optimizing selforganizing control architectures with genetic algorithms: The interaction between natural selection and ontogenesis. In Manner, R. & Manderick, B. (Eds.), *Proceedings of the Second Conference on Parallel Problem Solving from Nature* (pp. 451–460).

Armony J., Seruan-Schreiber J., Cohen J., & LeDoux J. (1995). An antomically constrained neural network model of fear conditioning. *Behavioral Neuroscience*, *109*, 246–257.

Baloch A., & Waxman A. (1991). Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot mavin. *Neural Networks*, *4*, 271–302.

Bienenstock E., Cooper L., & Munro P. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience, 2*, 32–48.

Dehaene S., & Changeux J.-P. (1997). A hierarchical neuronal network for planning behavior. *Proceedings of the National Academy of Sciences of the USA*, *94*, 13293–13298.

Denham, M., & McCabe, S. (1995). Robot control using temporal sequence learning. In *WCNN ’95. World Congress on Neural Networks*, Vol. 2 (pp. 393–396). Mahwah, NJ: Erlbaum.

Dickinson, A. (1994). Instrumental conditioning. In N. Mackintosh (Ed.), *Animal learning and cognition* (pp. 45-79). San Diego, CA: Academic Press.

Dominey P., Arbib M., & Joseph J. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience*, *7 (3)*, 311–336.

Douglas R., Mahowald M., & Mead C. (1995). Neuromorphic analogue VLSI. *Annual Review of Neuroscience*, *18*, 255–281.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Galuske, R., Singer, W., & Munk, M. (1997). Reticular activation facilitates use-dependent plasticity of orientation preference maps in the cat visual cortex. In *Society for Neuroscience Meeting*, New Orleans, LA, Abstracts, p. 2059.

Goldstein, L., & Smith, K. (1991). Bugworld a distributed environment for the study of multi-agent learning algorithms. Technical report, Department of Computer Science, UCSC.

Grossberg S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysical theory. *Psychological Review*, *89*, 529–572.

Grossberg S., & Levine D. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement. *Applied Optics*, *27*, 5015–5030.

Grossberg S., & Schmajuk N. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, *15*, 195–240.

Hall, G. (1994). Pavlovian conditioning: Laws of association. In N. Mackintosh (Ed.), *Animal learning and cognition* (pp. 15–43). San Diego, CA: Academic Press.

Hopfield J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, *79*, 2554–2558.

Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Indeveri, G., & Verschure, P. F. M. J. (1997). Autonomous vehicle guidance using analog VLSI neuromorphic sensors. In W. Gerstner, A. Germond, & M. H. Nicoud (Eds.) *Proceedings Artificial Neural Networks-ICANN97*, Lausanne, Switzerland (pp. 811–816). *Lecture Notes in Computer Science*. Berlin: Springer.

Kaelbling L., Littman M., & Moore A. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Klopf, A. (1982). *The hedonistic neuron: A theory of memory, learning and intelligence*. Washington, DC: Hemisphere.

König, P., & Verschure, P. (1995). Subcortical control of the synchronization of cortical activity: A model. In *Society for Neuroscience Meeting*, San Diego, CA. Abstracts, p. 1651.

Konorski, I. (1967). *Integrative activity of the brain*. Chicago, IL: University of Chicago Press.

Kröse, B. & Van Dam, J. (1997). Neural vehicles. In O. Omidvar & P. Van der Smagt (Eds.), *Neural systems for robotics* (pp. 271–296). New York: Academic Press.

Lavond, D. G., Kim, J. J. & Thompson, R. F. (1993). Mammalian brain substrates of aversive classical conditioning. *Annual Review of Psychology*, *44*, 317–342.

Mackintosh, N. (1972). *The psychology of animal learning*. New York: Academic Press.

Miller S., & Konorski J. (1928). Sur une forme particulière des reflexes conditionnels. *Comptes Rendus des Seances de la Societé Polonaise de Biologie*, *49*, 1155–1157.

Mondada, F., & Verschure, P. F. M. J. (1993). Modeling system–environment interaction: The complementary roles of simulations and real world artifacts. In *Proceedings of the Second European Conference on Artificial Life* (pp. 808–817). Cambridge, MA: MIT Press .

Mondada, F., Franzi, E., & Ienne, P. (1993). Mobile robot miniaturisation: a tool for investigation in control algorithms. In *Experimental Robotics III. Proceedings of the 3rd International Symposium on Experimental Robotics*, Kyoto, Japan (pp. 501–513). Berlin: Springer.

Moore, M. E. (1956). Gedanken-experiments on sequential machines. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies* (pp. 129–153). Princeton, NJ: Princeton University Press.

Morita M. (1996). Computational study on the neural mechanism of sequential pattern memory. *Cognitive Brain Research*, *5*, 137–146.

Newell A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135–183.

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

Oja E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford: Oxford University Press.

Rao R., & Ballard D. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*, 712–763.

Roberts, W. (1992). Foraging by rats on a radial maze:Learning, memory, and decision rules. In I. Gormezano & E. Wasserman (Eds.), *Learning and memory: The behavioral and biological substrates* (pp. 7–24). Hillsdale, NJ: Erlbaum.

Schultz W., Dayan P., & Montague P. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Sutton R. S., & Barto A. G. (1981). Toward a modern theory of adaptive networks: Expectations and prediction. *Psychological Review*, *88*, 135–170.

Thompson R. F., Berger T. W., & Madden I. V. (1983). Cellular processes of learning and memory in the mamalian CNS. *Annual Review of Neuroscience*, *6*, 447–491.

Thorndike, E. (1911). *Animal intelligence*. New York: Macmillan.

Touretzky, D. & Saksida, L. (1996). Skinnerbots. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollock & S. Wilson (Eds), *From Animals to Animats 4: Proceedings of the 4th International Conference on Simulation of Adaptive Behavior* (pp. 285–294). Cambridge, MA: MIT Press.

Verschure P. F. M. J. (1990). Smolensky's theory of mind. *Behavioral and Brain Sciences*, *13*, 407.

Verschure, P. F. M. J. (1992). Taking connectionism seriously: The vague promise of subsymbolism and an alternative. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, Bloomington, IN (pp. 653–658). Hillsdale, NJ: Erlbaum.

Verschure, P. F. M. J. (1993a). The cognitive development of an autonomous behaving artifact: The self-organization of categorization, sequencing, and chunking. In H. Cruze, H. Ritter, & J. Dean (Eds.), *Proceedings of Prerational Intelligence* (pp. 95–117). Bielefeld: ZiF.

Verschure P. F. M. J. (1993). Formal minds and biological brains. *IEEE Expert*, *8 (5)*, 66–75.

Verschure, P. F. M. J. (1996). Minds, brains, and robots: Explorations in distributed adaptive control. In A. Soares (Ed.), *Proceedings of the Second Brazilian International Conference on Cognitive Science*.

Verschure, P. F. M. J. (1997a). Connectionist explanation: Taking positions in the mind-brain dilemma. In G. Dorffner (Ed.), *Neural Networks and a New Artificial Intelligence* (pp. 133–188). London: Thompson.

Verschure, P. F. M. J. (1997b). Xmorph: A software tool for the synthesis and analysis of neural systems. Technical report, Institute of Neuroinformatics, ETH-UZ.

Verschure, P. F. M. J. (1998). Synthetic epistemology: The acquisition, retention, and expression of knowledge in natural and synthetic systems. In *Proceedings World Conference on Computational Intelligence 1998*, Anchorage, AL.

Verschure P. F. M. J., & Coolen A. C. C. (1991). Adaptive fields: distributed representations of classically conditioned associations. *Network*, *2*, 189–206.

Verschure, P. F. M. J., & König, P. (1997). Modulation of temporal interactions in cortical circuits. In H.-M. Gross (Ed.), *Proceedings of SOAVE97* (pp. 77–88). Dusseldorf: DVI.

Verschure, P. F. M. J., & Pfeifer, R. (1992). Categorization, representations, and the dynamics of system–environment interaction: A case study in autonomous systems. In J. A. Meyer, H. Roitblat & S. Wilson (Eds), *From Animals to Animats: Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior*. Honolulu, HI. Cambridge, MA: MIT Press, pp. 210–217.

Verschure P. F. M. J., Kröse B., & Pfeifer R. (1992). Distributed adaptive control: The self-organization of structured behavior. *Robotics and Autonomous Systems*, *9*, 181–196.

Verschure P. F. M. J., Wray J., Sporns O., Tononi G., & Edelman G. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems*, *16*, 247–265.

Weinberger N. M., Javid R., & Lepan B. (1993). Long term retention of learning-induced receptive field plasticity in the auditory cortex. *Proceedings of the National Academy of Sciences of the USA*, *90*, 2394–2398