Voegtlin, T. and Verschure, P. F. M. J. (1999) What can robots tell us about brains? A synthetic approach towards the study of learning and problem solving, *Reviews in the Neurosciences*, 10: 3-4, 291-310.

What can robots tell us about brains? A synthetic approach towards the study of learning and problem solving.

Thomas Voegtlin, Paul F.M.J. Verschure,
Institute of Neuroinformatics, ETH-UZ,
Zürich, Switzerland.
thomas@ini.phys.ethz.ch, pfmjv@ini.phys.ethz.ch

Keywords

Classical conditioning, operant conditioning, short-term memory, long-term memory, replay, chaining, locality, sequence learning.

Corresponding author:
Paul F.M.J. Verschure,
Institute of Neuroinformatics, ETH-UZ
Winterthurerstrasse 190, Zurich,
CH 8057, Switzerland.
phone : +41 1 635 30 70
fax : +41 1 635 30 53

Synopsis

This paper argues for the development of synthetic approaches towards the study of brain and behavior as a complement to the more traditional empirical mode of research. As an example we present our own work on learning and problem solving which relates to the behavioral paradigms of classical and operant conditioning. We define the concept of learning in the context of behavior and lay out the basic methodological requirements a model needs to satisfy, which includes evaluations using robots. In addition, we define a number of design principles neuronal models should obey to be considered relevant. We present in detail the construction of a neural model of short- and long-term memory which can be applied to an artificial behaving system. The presented model (DAC4) provides a novel self-consistent implementation of these processes, which satisfies our principles. This model will be interpreted towards the present understanding of the neuronal substrate of memory.

1 Introduction

The systematic investigation of animal learning and problem solving started about one hundred years ago with the work of Thorndike and Pavlov [44, 33]. These studies introduced two paradigms which have since then dominated the field; operant and classical conditioning. Operant, or instrumental, conditioning describes tasks where animals learn on the basis of the consequences of their own actions. Thorndike used a, so called, "puzzle box" (Figure 1.A), where an animal, a cat or dog, had to learn a specific sequence of actions in order to escape from the box. Using these examples of trial and error learning Thorndike showed that performance, as measured by time to escape, improved over trials. The paradigm of classical, or Pavlovian, conditioning refers to learning phenomena where initially neutral, *conditioned stimuli* (CS), such as lights and bells, become through their correlated presentation with motivational, *unconditioned stimuli* (US), like footshocks or food, able to trigger a *conditioned response* (CR). In the early work of Pavlov this involved the induction of conditioned salivation (CR) to a bell (CS), using food as an unconditioned stimulus (Figure 1.B).

Insert figure 1 about here

Thorndike's research is an early example of comparative psychology, where the differences between human and animal problem solving were investigated. Thorndike's goal was to place this line of research on a firm empirical footing as opposed to the more anecdotal approach of his predecessors (i.e. [37]). He aimed at isolating the laws that govern the learning process. His most famous proposal is the so called *Law of Effect*, which states that

associations develop according to the outcome of actions; rewarded actions strengthen associations while punished actions weaken associations. In case of Pavlov the focus was on the neuronal mechanisms underlying the forms of learning he initially observed while investigating the digestive system. Both influential paradigms have over the last century led to an extended program of research in psychology, ethology, and neuroscience. They have also formed the driving force behind the behaviorist revolution of the twenties and thirties, with its emphasis on a strictly empirical approach towards the study of behavior. The restriction to "observables" imposed by this approach[1], however, together with the development of computing machinery induced a shift to a more integrative, multidisciplinary approach, cognitive science [12]. The aim of cognitive science was to open the black box which intervened between the stimuli and responses manipulated by the behaviorists.

Today, the study of mind, brain, and behavior is a strongly multidisciplinary field, also known as cognitive neuroscience. Properties of the brain and behavior are described over a wide range of levels: from molecules, ion channels and cells to circuits and systems. These different levels of description have been progressively investigated by more and more researchers, who have become increasingly specialized. The collective database of their efforts has taken on enormous proportions. An immediate consequence of this specialization is an unprecedented fragmentation of knowledge which can be seen as one of the main limiting factors in our understanding of mind, brain, and behavior. This problem is not unique for this domain.

[1] although it needs to be emphasized that this is certainly not true for many researchers of this period (e.g. [16])

Similar observations have been made in biology, where Strohman [41] interprets this as a sign of a scientific crisis, and in psychology, where Newell [31], identifies the "great psychological data puzzle" and proposes that a synthetic approach, artificial intelligence, can alleviate this situation.

As an illustration of the fragmentation of scientific knowledge we can consider the issue of learning, the subject of our own studies. The concept of learning traditionally designates long-term changes in the behavior of a system. Psychologists have accumulated a large amount of observations on the behavioral regularities that can be observed under specific, often rather artificial, conditions over a wide range of animal species, from snail to man [23, 11]. Neuroscientists have added to this set observations on effects which are dependent on particular lesions or pharmacological challenges. Alternatively they have reported on correlations between physiological measures and performance [24]. Other neuroscientists have investigated the subcellular changes associated with learning, for instance using the popular paradigms of long term potentiation and depression [2]. These investigations are often based on the common assumption that the substrate of learning is provided by synaptic plasticity. Others, however, would argue that the neuronal substrate of learning needs to include more general changes in neuronal morphology and interconnectivity patterns (e.g. [14]). These approaches are further complemented with explorations at a genetic level [5]. At the macroscopic level of complete systems novel imaging techniques have opened up a window on the processes involved in learning and memory in the human brain [40]. The above demonstrates the wealth of methods and

techniques. The guiding principle of how these are employed, however, is in general to detect a correlation between a particular manipulation of the behaving system and brain derived measures. An added complicating factor in such an approach is that not only differences between species, but also for instance between strains, gender, age, and the circadian rhythm need to be considered [1, 10]. Given the tremendous advances in the technologies available the space of possible correlations must be considered practically infinite. Given this wide range of perspectives on learning, the question can be raised, whether the same phenomenon is studied in all these approaches. Although the cognitive revolution might have opened the black box, the pieces presently appear to us in a highly disordered manner. The need for a blueprint of the underlying design principles is evident.

We do not want to claim that no proposals are available on the principles of behavioral and neural organization, which underly the phenomena described in the collective neuroscientific database. For instance, in the case of classical conditioning the model of Rescorla and Wagner [36] (see [27] for a review), provides a good description of many behavioral regularities observed in this learning paradigm. The basic assumption behind this model is that the effect of reinforcement, derived from a US, on the association between a stimulus (CS) with the unconditioned response is not only dependent on the properties of that particular stimulus but also upon the properties of the other stimuli known to the system; learning is based on the violation of expectations. The model aimed specifically at accounting for the phenomena of blocking and overshadowing [17, 18], which demon-

strated that learning does not seem to follow Thorndike's Law of Effect, but depends on "previous knowledge" of the organism. Although this model has in turn been criticized on various grounds (see [23, 11, 48]) it makes accurate predictions on the behavioral changes which can be observed in classical conditioning.

Given the overwhelming amount of data, and the relative lack of hypothesis on underlying principles, we need to consider whether a pure empirical investigation of the phenomenon of learning, or any other construct applied to neuronal function for that matter, will help us to understand the basic principles of neuronal organization, which find their expression in this myriad collection of research paradigms. There is no reason to admit defeat, but this situation can be taken as a challenge to reconsider the basic approaches followed. In this paper we want to demonstrate how a synthetic approach can provide a research strategy which is complementary to the empirical mode of research, common in the brain and behavioral sciences. A synthetic approach, for example using computer simulations, can facilitate the development and exploration of scenarios on the principles of neuronal organization. Before elaborating on the methodological considerations behind such a proposal we want to further define the concept of learning.

Following earlier proposals [35] we assume that behavior serves to guarantee the integrity of the behaving system [46]. In the context of this assumption we propose that learning is a response of biological systems to a certain type of unpredictability [45]. Indeed, the genomic plan of an organis-

m has to address two types of unpredictability: somatic and environmental. Somatic unpredictability results from the various ways the body plan can be realized, depending on the highly nonlinear and complex interactions between the genes, the phenotype and the environment. Environmental unpredictability means that biological systems, specifically vertebrates, will have to deal with an environment whose crucial properties are *a priori* unknown. Despite this uncertainty they succeed in performing a wide variety of tasks. The knowledge required to accomplish these tasks can be acquired, essentially because the world has some regularities that can be *learned*. Hence, we call *learning* any structural change to a behaving system, that captures regularities of its interaction with an environment that were not predicted by its genome, as to allow these regularities to be exploited in its behavior. Biological systems that express learning are able to deal with a wider range of tasks and environments than systems that do not. The paradigms of classical and operant conditioning reflect adaptations to conditioned stimuli that can be *a priori* of any kind (they are only constrained by the properties of the sensors) and they illustrate this versatility.

A synthetic approach is based on the construction of models. Given present day computer technology we have the unique opportunity to realize thought experiments on scenarios representing principles of neural organization. These realized thought experiments, however, acquire scientific meaning only through their interaction with the domain of empirical observation. It is important to consider in more detail the methodological considerations behind a synthetic approach. On one hand, the aim of a model needs to

be considered. Models allow us to summarize large numbers of observations on a certain phenomenon in a rather concise way in terms of assumed underlying variables and parameters. This facilitates communication and evaluation. On the other hand we need to be concerned with the validity of a model. In general a model tries to describe a certain input-output relationship, *response function*, in terms of a *transfer function f*: output = *f*(input). The observations which express the input-output relationship will consist of a number of points in some multidimensional space. A model can be seen as a means to draw a continuous line through these points. As an example we can consider the model of Rescorla and Wagner, discussed earlier, which makes predictions on learning curves, which are measured in terms of the fraction of observed CRs after a certain number of learning trials. These types of *descriptive models*, however, are confronted with a fundamental problem. In principle an infinite number of lines can be drawn through the observed response function. This problem of *indeterminancy* was first pointed out by Moore in 1956 [28].

The only way to answer this challenge is by imposing additional constraints on the set of possible transfer functions. However these additional constraints are taken from other levels of description; *convergent validation* [47]. This implies, however, that a model needs to be defined as a *generative model* where the transfer function becomes a macroscopic variable of the defined system, while its central parameters are defined at its microscopic level. As an example we can consider the influential model of Hodgkin and Huxley [15], which describes how the macroscopic property of axons to ini-

tiate and propagate action potentials can be accounted for in terms of the interaction of a number of microscopic components; a sodium, potassium, and leak conductance, which change depending on the ion concentration and an electrical gradient. Hence, in order to address the problem of indeterminancy, models should necessarily be required to be generative, satisfying constraints from multiple levels of description; i.e. anatomy, physiology, and behavior. The combination of our conceptualization of learning and these methodological considerations constitutes a program of *synthetic epistemology* [54]; the study of learning by biological systems following a multilevel synthetic approach based on large scale computer simulations and real-world devices; robots.

1.1 Robots

Is there a difference between a brain, a robot and a computer? What we call a robot is an artificial behaving system that can interact with an environment. There is no reason to believe that natural brains are intrinsically able to perform operations unaccessible to computers. But our previous definition implies that *learning* is possible in natural or artificial systems only if they interact with an environment. Hence, models that include robotic components can approach the study of the principles of neural organization in a more powerful way than methods that restrict themselves to internal computations, since they can account for the various interactions between a behaving system and its environment. The "knowledge" developed by a behaving system (natural or artificial) through a learning process depends first on the properties of its control structure. However, another limiting factor is the complexity of its environment, which generates the stimuli. Since

learning implies that some regularities exist in the world, complex learned abilities need, in order to emerge, a world with complex properties.

Robots can be real-world devices, but it is also possible to simulate behaving agents and their environment using computer programs. In our research we use both approaches [30]. Using real-world devices can ensure that the complexity of the environment will not be a limiting factor of learning. However, simulated robots allow a systematic evaluation of *all* the parameters that are relevant for the learning process, and guarantee repeatability of the experiments.

In our further analysis, we will present our work on learning and problem solving as an example of a synthetic approach based on the above methodological and conceptual considerations. Since the aim of the present paper is to provide an illustration of the potential of this approach we will focus on describing relevant examples from our own work. In particular, we will describe in more detail the development of a fully neurally realistic system of short and long-term memory which is evaluated in the context of artificial behaving systems. This serves to illustrate the different aspects of a synthetic multilevel approach towards the study of mind, brain, and behavior. Given these aims we will not provide an exhaustive comparison with the existing literature relating to the details of the presented models.

1.2 The learning hypothesis

In order to explain the forms of learning revealed through the experimental paradigms of classical and operant conditioning, we assume that they can be described by different, but interacting, levels of control. First, unconditioned responses can be derived from a *reactive control* structure. This structure implements prewired relationships between US events and URs, and will reflexively respond to immediate events. Since the set of unconditioned stimuli is derived from genomic information, these stimuli must be simple and based on low complexity sensors, in general proximity sensors. Unconditioned responses reflect actions of a behaving system in response to *specific* events. For instance, a burning hot contact on the hand triggers a contraction of the arm. Reactive control provides the behaving system with a basic level of competence to deal with its environment and prevents its disintegration.

Second, the tuning of the responses of an organism to *non-specific* events can be accounted for by an *adaptive control* structure. Since non-specific events are *a priori* unknown, this structure will need to develop representations of events that are relevant (the CS). The criterion of relevance is the correlation of CS events with unconditioned stimuli, or previously acquired conditioned stimuli. The representation of CS events is *constructed* at the level of adaptive control. This level of control approximates relations between CS and US events through instantaneous correlative measures, and triggers conditioned responses to conditioned stimuli. At the level of an adaptive control structure the detailed properties of a UR, such as its onset

and duration, can be changed to create a CR (specific learning).

Third, correlations between stimuli that are not instantaneous can be captured by forming sequential representations of sensorimotor events. A level of control forming sequential representations (*contextual control*) allows the behaving system to acquire "plans" involving its future actions and the expected stimuli resulting from these. For instance, in Thorndike's puzzle box, a cat had to perform several actions in a precise order, for it to escape from the box.

Our hypothesis is that these three levels of control are sufficient to account for both classical and operant conditioning phenomena. Distributed Adaptive Control (DAC) are a series of models that implement these three levels of control using artificial neural networks. They are evaluated in the control of behavior using robots [49, 50, 46, 30, 51, 53].

1.3 Principles of neural design

Given our incomplete knowledge of the biological mechanisms of learning and problem solving, it is necessary to constrain our choices of implementation. In this case we want to particularly emphasize the constraints imposed on information transfer in biological systems. A neuron can only use the information that is locally available, through synapses or other forms of chemical transmission. In particular, it is not possible to move a pattern of activity from one population of neurons to another using a supervisor that would pick the information somewhere in the network and move it to anoth-

er place. This constitutes a *principle of locality*. This principle is true for space (*spatial locality*) but also holds for time; if a pattern of neural activity has not changed the structural properties of the substrate, (e.g. synapses, cell morphology), it cannot be reconstructed later (*temporal locality*). It is fundamental to respect these principles in the design of control structures, if one doesn't want to violate the obvious facts known about biology. A third principle guiding model development is to minimize the complexity of the network. This is not only based on common sense (Ockham's razor), but also on the observation that in case the testable components of a model are provided by its assumptions, starting a model based on a super-powerful description method would preclude any further validation [25].

2 Methods

The behavioral task we use to study our models of control is a foraging task, where an agent has to avoid collisions with obstacles while locating targets dispersed in its environment. Experiments are performed either in a simulation environment, BugWorld [13], or using a real-world robot (Khepera, K-team,Lausanne) with the IQR421 distributed simulation environment [52].

Insert figure 2 about here

BugWorld is a two-dimensional environment containing obstacles, targets, and circular robots. The body of a simulated robot is called the *soma* (figure 2.A). BugWorld robots have proximal and distal sensors. Their distal sensors respond to the distance to surfaces in their field of view. The

proximal sensors are target and collision sensors. The target sensors are placed at 90^o and -90^o from the axis of the soma. They detect a signal emitted by the targets, which is a decreasing function of the distance to the targets. For the Khepera robot (figure 2.B), the targets are light sources. The proximal sensors of the Khepera robot are infrared (IR) sensors, with which the immediate proximity of IR reflecting surfaces can be detected, or ambient light levels can be measured. Its distal sensor is a color CCD camera.

In DAC, proximal sensors generate unconditioned stimuli (US) while distal sensors generate conditioned stimuli (CS). The unconditioned stimuli can be of two types: aversive (US-) or appetitive (US+). Appetitive stimuli come from the targets and the associated reflexes are *approach actions*. Aversive stimuli are collisions with obstacles, and the associated reflexes are *avoidance* of the obstacle.

3 The Distributed Adaptive Control series

3.1 DAC0: The reactive control structure

DAC0 is our implementation of a *reactive control structure*. It is fully prewired and its control consists of basic reflexes or stereotypic behavioral patterns.

The control architecture DAC0 consists of 3 types of neurons[2] (figure 3):

- Internal state units (IS) receive inputs from the US sensors. They can be of two types: aversive ($IS-$) or appetitive ($IS+$). The $IS-$

[2]What we mean with neuron is an approximation of a biological neuron, that sums its inputs and gives an output value which is a nonlinear function of this sum.

group gets inputs from the collision sensors while the $IS+$ group gets inputs from the target sensors. The IS cells are active when the corresponding collision sensor element is activated.

- Actions are triggered by a group of motor units (UR). UR receives its inputs from the IS cells. The inputs received from $IS+$ trigger approach actions while the inputs from $IS-$ trigger avoidance actions.

- An inhibitory unit I is excited by aversive events ($IS-$) and inhibits the appetitive cells $IS+$. This provides the agent with priorities between approach and avoidance behaviors; conflict resolution.

Insert figure 3 about here

A trajectory of DAC0 consists of typical events (figure 4). The behaving agent can move forward, turn to the right or to the left. In the absence of any stimulus, it moves forward, which constitutes exploration. Starting at position 0, DAC0 explores its environment (translational movements). In positions 1,2 it collides with obstacles and each collision induces a turn to the left (avoidance action). At location 3 the target A is detected and an approach behavior is induced. Another collision occurs at location 4, triggering a turn to the left. In location 5, the soma follows the gradient of the signal until the target is found.

Insert figure 4 about here

3.2 DAC2: Adaptive control structure

The *adaptive control structure*, DAC2, learns to correlate CS events (distal sensor) with internal states (IS). It is an implementation of the non-specific

component of classical conditioning. DAC2 includes the reactive control of DAC0. Initially, the behavior of DAC2 is entirely made up of the unconditioned reflexes triggered by its reactive control structure. This reactive structure constrains any subsequent learning process.

Insert figure 5 about here

We propose that a central element of classical conditioning is *CS identification*. Thus, DAC2 has another population of units, CS, which receive their inputs from the distal sensors (figure 5). Learning at the adaptive level consists in "categorizing" the CS events and classifying their correlations with US events. Categorization means that a prototypical representation of the CS is constructed from the input CS. Learning leads to the control of the UR cells by the CS population; In case a relevant CS event is recognized, the activity of the CS cells is propagated to the IS units which in turn activate the motor units through the predefined connections between IS and UR.

Learning the connections between CS and IS cells is based on a *reconstruction*: First, excitatory connections from CS to IS translate the activity of CS into a pattern of activity in IS. Then, inhibitory feedback connections from IS to CS propagate a *prototype* of the CS, dependent on IS activity, which is subtracted from the activity of the CS cells. The difference between the actual CS and the CS prototype is called *reconstruction error*. The modifications of the symmetric synaptic weights are proportional to this error.

The activity v_i of unit i in the IS population is:

$$v_i = \sum_j w_{ij} u_j + c_i \tag{1}$$

where c_i is the component that depends on the US, u_j is the activity of unit j in CS, and w_{ij} is the synaptic weight between i and j. The IS population in turn inhibits the CS population, generating a prototype. The prototype vector p is defined by:

$$\forall j, p_j = \sum_j w_{ij} v_i \tag{2}$$

where p_j is the predicted activity of CS unit j given the activity in IS. After this feedback, the activity of cell j of the CS population, u'_j, is defined as $u'_j = u_j - p_j$, which corresponds to the reconstruction error. The weights of the connections between CS and IS are updated according to a Hebbian learning rule:

$$\forall i, j, \Delta w_{ij} = \eta v_i u'_j \tag{3}$$

where η is a learning rate.

This learning rule is defined on the basis of a number of observations derived from our robotic experiments. In [49] it was shown that in order to acquire and retain CS-US associations in a behaving device a correlation based learning rule needs to include an activity dependent depression term. This renders a learning rule equivalent to the, so called, Oja learning rule [32]. It was demonstrated, however, that this solution becomes unstable over long periods of time. The observed instability of this local learning rule, primacy and overgeneralization, was solved by embedding the process regu-

lating synaptic efficacy in a recurrent circuit [50], and was further developed in [53].

Figure 6 shows the representations of CS events expressed in the strength of the synapses between the CS and IS populations of a real-world agent. The environment of figure 6.A has regular properties; different US events are correlated with the presence of patches of different colors that are detected by the visual system of the robot. This system uses 36 cells, 12 for each color (red, green and blue). Each cell covers a unique 45x30 pixels region in the 640x480 image from the camera (see [53] for details). In this environment, the robot learned to associate particular colors with particular US events. Figure 6.B displays the time evolution of the synaptic weights of the adaptive control structure, after 1, 1.5 and 2 hours. Not only are the correlations present in the environment accurately reflected in the interconnectivity, but individual cells in $IS-$ and $IS+$ develop unique representations of particular collision or target events. For instance, the white rectangle in the second row, first column of the "red – $IS-$" matrix shows that the collision detector number 2 was correlated with the presence of red in visual region number 1 (upper left corner), resulting in a high synaptic weight. Comparing the three displays we observe that the invariants extracted from the environment by DAC2 remain stable over an extended period of time (figure 6.B).

Insert figure 6 about here

Learning in this control structure is, however, limited to immediate correlations between CS and US events. The result is that the behavior of DAC2 entirely depends on current sensory inputs. The agent can explore its environment and extract some general properties, but cannot learn tem-

poral relations between multiple events.

3.3 DAC3: A contextual control structure

The aim of the third level of control is to allow the *acquisition* of sequential representations of events, to *retain* them in a memory, and to *express* them in behavior. In a task of sequence learning, the response (output) of a system does not only depend on the immediate input, but also on the context provided by previous inputs; *temporal context* [56]. An agent provided with this third level of control is able to choose its actions based on both the *temporal context* and on its experience. This level is called *contextual control*.

A sequence consists of sensorimotor events; *segments*. A segment is a couple consisting of a CS prototype, constructed by the adaptive control structure (eq. 2), and an associated motor action (UR).

A contextual control structure will need to select certain sensorimotor sequences, among the whole set of behaviors generated by the adaptive control structure. Sequences that need to be selected for acquisition are those that lead to a modification of an internal state. For instance, in our foraging task, we use the contextual control structure in order to find targets. In this specific task, a sequence of actions that leads to a target is a *rewarding sequence*.

Insert figure 7 about here

In order to acquire a sequence, it is necessary to remember the sensorimotor events that have preceded the modification of the internal state.

Since we do not know what the outcomes of our actions are, it is necessary to have a mechanism that continuously stores events and is able to retain them; *short-term memory*. It should be emphasized that since we have no *a priori* knowledge of what events will later trigger a change in the internal state, like the delivery of a reward, the short-term memory needs to keep track of *any event* at any time.

In order to modify the behavior, sequences that preceded a modification of an internal state have to be stored in a selective memory that keeps track of these events over a longer period of time than the short-term memory. For the present discussion, we refer to this component as *long-term memory*. This definition is more restrictive than the definition of long-term memory generally used in psychology, which designates all long-term changes [42]. For instance, learning at the DAC2 level is a form of long-term memory but for our present discussion it is considered as a separate mechanism. While the agent explores its environment it compares its sensory inputs with the content of its long-term memory in order to use its learned behaviors. If the current CS prototype and the current context match a learned situation, then the agent executes the corresponding motor action stored in its long-term memory.

In our foraging task, rewarding sequences are acquired during *stimulation periods* where targets emit a signal. The expression of learned behaviors can be observed during *recall periods* where the signal emitted by the targets has been suppressed.

For our implementation we make an additional distinction between *default* and *non-default* actions. During the stimulation periods, the signal from the targets (US+) can trigger approach actions of the adaptive structure, or suppress avoidance actions, if the influences of the US+ and the US- are balanced. Actions that depend on the US+ are called *non-default actions*. If no US+ is detected, *default actions* are generated by the adaptive control structure. Since we want to use the contextual control structure in order to find targets during recall periods, only the non-default actions need to be considered in the sequence learning task.

DAC3 is our first implementation of an agent with contextual control [46]. Its contextual control structure is built on top of the same adaptive control structure as DAC2. The short-term memory of DAC3 is a ring buffer that stores the last sensorimotor events. Each time a target triggers a non-default action, a CS-UR couple is stored in this buffer. The stored CS event corresponds to the prototype that has been derived from the stimulus, the UR event corresponds to the triggered motor action. If a target is found then the sequence contained in the short-term memory is stored in the long-term memory. Hence, the long-term memory is a list of sequences of sensorimotor events.

During exploration the actual CS prototype (equation 2) is compared to the prototypes in the segments of the long-term memory. This comparison is followed by a selection; the best-matching unit (winner), if its prototype

is close enough to the actual CS prototype, induces an action by activating the UR units. This selection, however, is biased since the winner unit will enhance the likelihood that the next segment of its sequence will win the competition in the future. This bias allows the actions of the agent to be dependent on context (*chaining*).

Preliminary experiments showed that DAC3 is able to display structured behaviors, such as stable trajectories between targets [46]. In [53], we have shown that the contextual control structure allows DAC3 to find more targets than DAC2 during recall periods, when the signal from the targets is suppressed.

4 DAC4: A neural implementation of a contextual control structure

The principles underlying the contextual control structure of DAC3 are competition between simple units and selection. These mechanisms are fundamental principles in unsupervised training of neural networks, and their role has been considered in natural systems, [4, 21]. So far, however, for the short-term and long-term memory structures of DAC3 we made use of ring buffers and chained lists. In this way the implementational issues were side-stepped in order to investigate the functional properties of a contextual control structure. These preliminary investigations established that sequential learning could be explained in these terms [53]. The question whether the same functional properties could be implemented in a biologically plausible way raises some important challenges.

DAC4 is our first fully neural implementation of a contextual control structure. It is consistent with the principles of locality presented in the introduction; it does not violate the obvious knowledge that we have of biological processes of learning. Addressing the issue of the neural design of this control structure allows the investigation of the constraints imposed on natural nervous systems, in terms of acquisition, retention and expression of information. These are fundamental questions that cannot be addressed by the classical approach of designing neural networks that perform isolated tasks, given the relationship between a control structure, the properties of the soma and of the environment we showed earlier.

4.1 Requirements

Before we present the model we want to specify in detail some functional requirements of a neural structure that acquires, retains and expresses sequential information.

4.1.1 Sequence learning with ANNs

A contextual level has to acquire sequences of sensorimotor associations. Sequence learning means that the responses generated by the network are more than simple associations between inputs and outputs, but also depend on the temporal context provided by its previous inputs [56].

Sequence learning with neural networks has been investigated in various ways [9], [34], [19]. A robust class of methods use networks with recurrent connections, so that the pattern of activity of the cells in the recurrent

loop depends on the temporal context of past events, thus having units representing context [8]. Most of these models, however, use the supervised backpropagation of an error signal in their learning rule, and this error term contains nonlocal information. In addition these models face difficulties in the representation of temporal contexts with long-term dependencies [3].

(Dominey et al, 1995) showed that sequence learning is also possible with a recurrent network, using only local information [7]. This network is made of two interconnected populations, *State* and *Context*, and an output population. The synaptic weights of the recurrent connections between *State* and *Context* are randomly chosen, and not plastic. Sequences are learned using a Hebbian learning rule between units of *State* and units of the output layer (associative memory). In this case, the temporal context is represented by the pattern of activity in *State* and *Context*; the activities of the cells in these populations depend on the temporal context. This representation is *predefined* by the random connections. This network has been applied to the study of corticostriatal plasticity, and the dynamics of prefrontal cortex [7, 6]. Since it does not violate the principles of locality and has the robustness of recurrent networks, we tried to adapt it to our task of sequence learning.

4.1.2 The short-term memory

Can we use a recurrent model like the one presented in [7] for the short-term memory? This model is in general not able to learn a sequence using one single presentation, because it uses an associative memory. A short-term

memory, however, needs to acquire sequences immediately. Consider the case of a network that needs several presentations of a sequence in order to successfully store it. In order to function as a short-term memory, such a network would need to acquire sequences at the moment when they are presented by the external environment, and not at moments that depend on internal constraints imposed by the network. For instance, if two presentations of a rewarding sequence are separated by an interval of one day, then this network would have to maintain the preliminary sketch of learning for one day before it could be refined. However, as we discussed earlier, the short-term memory needs to keep track of *any event*. Therefore, all the events that happen within this day would have to be acquired in the same way, without erasing the first preliminary sketch of learning. Since the outcome of actions cannot be known in advance, such a memory would have to store an excessive amount of information and thus need a gigantic capacity. Therefore, according to the definitions of short-term and long-term memories given in section 3.3, it is necessary that the short-term memory acquires potential rewarding sequences after one single presentation. In this case, its content can be retrieved for long-term storage when there is a modification of an internal state, or erased during foraging.

4.1.3 The long-term memory

Can we use the same recurrent neural network for the short-term memory and for the long-term memory? In the model of (Dominey et al, 1995) [7] the representation of context is *predefined* by the random connections. According to our definition, the task of a short-term memory is limited to storage

and restitution of sensorimotor sequences. For this a predefined representation of context is not a problem, as long as the sequence of events can be retrieved. However, a long-term memory has to do more than storage and retrieval of information. As a physical system, it will have a finite capacity. However, since it will have to learn a virtually infinite set of sequences, a long-term memory needs to build *categories* and perform *generalization*. A predefined representation of temporal context cannot perform generalization and would restrict the set of possible categories. This means that the representation of context used by a long-term memory needs to adapt itself to the data rather than be fixed.

Another requirement of long-term memory can be called *identifiability*: In the comparison of current events with the content of the long-term memory, all the elements of any sequence are potentially relevant. Thus, they all need to be accessible at every moment. This favors representations of distinct prototypes by distinct units, instead of complex patterns that are attractors of the dynamics of a recurrent networks [53].

4.2 The model

Given the above considerations, we choose to implement the long-term memory and the short-term memory structures using two different neural networks. These separate networks satisfy the above requirements. The following sections are a general presentation of these networks. For a complete description, see the appendix.

4.2.1 The short-term memory

Most models of sequence learning with a neural network need several presentations of a sequence in order to learn it. But as argued above the short-term memory of a contextual control needs to acquire sequences after the first presentation.

Insert figure 8 about here

In order to solve this problem, we modified the model of Dominey [7], adding a new population of cells, called *Segments* (see figure 8). The populations of the recurrent network for the short-term memory are called *State* and *Context*. Units of *Segments* get inputs from the *State* population and associate *State* activities to sensory prototypes (in the CS population) and motor actions (in the UR population). At each time step, a competition selects a new unit of *Segments* which learns the association between the current pattern of activity in *State*, resulting from past events, and the current pattern of activity in CS and UR. The cells in *Segments* have extremely plastic synapses; selected cells learn the association immediately. The counterpart of this plasticity is that the information retained in the *Segments* population might be erased quickly when new associations are formed. We use a competition mechanism that favors units which have not been selected for a long time, in order to prevent quick overwriting of acquired associations. What prevents sequences to be forgotten on a longer time scale is their retention in long-term memory.

4.2.2 Transfer from short-term memory to long-term memory: Retention, Replay

The choice of having two separate populations of cells implementing the short-term and long-term memories implies that the information acquired by the short-term memory needs to be retained in the long-term memory; *physically moved* to another structure. In order to do this, our principle of spatial locality allows one possibility, which is to reactivate the sensory and motor cells corresponding to a sequence in order to modify the long-term memory synapses. It is during this "replay" that sequences are stored in the long-term memory.

This raises another question: How will sequences be replayed? Sequences are made of sensorimotor events, but it is not obvious whether one needs to replay them in the order of acquisition or not. They could also be replayed in a reverse order, or in a random order. We choose to replay events in the same order as they were acquired, because the long-term memory relies on a recurrent representation of temporal context which reflects the order of events. The other possibilities, however, cannot be excluded *a priori* (see discussion).

In order to replay a sequence acquired by the short-term memory, a first unit of *Segments* needs to be selected (figure 8), which initiates the replay. The selected cell does not change its plasticity, but excites cells in the sensory population corresponding to the CS and the associated motor units. Motor actions are inhibited during this phase. The sensory activation is propagated in the recurrent network of the short-term memory, *State* and

Context, which triggers a representation of the context corresponding to the next sensorimotor couple of the sequence, and to the selection of the next corresponding *Segments* unit. This loop allows to replay events in their order of acquisition. The result of this replay is the reactivation of sensorimotor units in the order of the sequence.

4.2.3 Initiating the replay

The choice of replaying sequences leads to an additional problem: Our recurrent network is able to retrieve a sequence starting from its beginning, or from any point of the sequence, but needs to be put in the state of activity corresponding to the starting point of the replayed sequence. However, at the moment the replay is initiated, the previous patterns of activity of the network are lost. Given the constraint of temporal locality, the short-term memory has to "retrieve" this starting point by translating a set of synapses into a pattern of activity. In addition, the "starting point" of the rewarding sequence is *a priori* not defined for the network. Thus, how shall the replay be initiated?

A possibility could be to use an additional system that acts as a supervisor. This system would store 'salient' events in order to use them later as starting points for the replay. This solution would add complexity to the network, and raise multiple problems such as: "what are salient events?", "when should they be forgotten?", etc. For this reason, we did not use this option. Our solution consists in adding a random perturbation to the activity of the cells in *State* and *Context*, in order to select the unit in *Segments*

which initiates the replay. Since this pattern of activity results from a random perturbation, it is not sure yet whether the network will replay events in the order of the acquired sequence. The evolution of a perturbation of a recurrent network is linked to properties of its internal connections, of the responses of neurons, and to the time constants used. In particular, there is a set of conditions for which the stored sequences are attractors of the dynamics (such a system is called Lyapunov-stable). In this case, the amplitude of the perturbation decreases during the replay, allowing for a replay of the segments in their order of acquisition. We use such a set of conditions for the recurrent network made of *State* and *Context*. Thus, during the replay process a learned sequence will be retrieved, which is an attractor of the dynamics.

This does not guarantee that the replayed sequence will correspond exactly to the events that led to the reward. However, depending on the number of units in *Segments*, and on the amplitude of the random perturbation, one can influence the probability to replay the relevant events.

4.2.4 The long-term memory; retention, expression

Insert figure 9 about here

We mentioned the need to have an adaptive representation of context in the long-term memory.

The layers of the recurrent network used for the long-term memory are $STATE$ and $CONTEXT$ (see figure 9). Instead of having predefined connections, as in [7], units in $STATE$ learn to respond to the coincidence of a

sensory stimulus and a context represented in $CONTEXT$. The activity in $CONTEXT$ depends only on the previous activities in $STATE$. (A detailed description of the learning rule used is provided in the appendix). Units in $STATE$ learn to efficiently represent sequences that are often presented, and are less able to represent situations that are rarely present in the exploration task. On each time step, the unit of $STATE$ which has the highest response activates the motor units, if its response is above a given threshold.

This principle allows a high flexibility in the execution of sequences; the cells of $STATE$ respond to the stimulus and also to the state of advancement of the behavioral plan that has been started. This allows to adapt the behavior of the agent when events in the world do not correspond to the learned sequence.

Not all the components of a sequence have to initiate an action, but the representation of context has to be maintained at all time, in order to allow the continuity of the executed plan. As in DAC3, we make the distinction between default and non-default actions. In order to reduce the sequence learning task to the necessary non-default actions, the representation of temporal context in $CONTEXT$ depends on the nature of the current action: Between two non-default actions, the activity of the cells of the $CONTEXT$ layer slowly decays. This makes the cells in $STATE$ learn how much time steps have elapsed between the sensory events corresponding to non-default actions. So the system also learns to let the same time elapse when the sequence has to be executed.

Unlike the short-term memory, this recurrent network will need several presentations of a sequence in order to learn it. For the long-term memory, this is not a problem since it is possible to replay the same sequence several times. Alternatively, the agent may have to find the target several times if the sequence is replayed only once each time a reward is found.

Insert figure 10 about here

Figure 10 shows an example of a successfully learned sequence. In a first presentation (10.1) the signal from the target is detected. When the signal is removed (10.2) the agent does not find the target anymore. After several presentations and replays of the sequence, the target can be found without the signal (10.3). Note that in 10.3 the motor sequence is not exactly the same as in 10.1. They can be made identical with further presentations. More generally, experiments showed that the behavior of DAC4 is very similar to the behavior of DAC3; both are implementations of the same contextual control. This demonstrates that the principles of contextual control explored by DAC3 can be implemented in a consistent way obeying the principles of spatial and temporal locality.

5 Discussion

This paper aimed at conveying a need to find approaches which can help us to explore principles of neural organization. We propose that synthetic methods, for instance based on digital simulation, provide an example of such an approach which is complementary to the more traditional em-

pirical mode of research in the study of brain and behavior. A synthetic approach, however, needs to follow a methodology which we summarized under the notion of convergent validation. This means that models need to satisfy constraints taken from multiple levels of description. As an example of such an approach we have reviewed our own work on learning and problem solving, Distributed Adaptive Control. In this context, learning is studied from a perspective that includes the environment, the phenotype, and the detailed properties of its control structure (brain). We consider our own work as providing a theoretical framework which at this point in time is self-consistent, it obeys the principles of locality and connects principles on physical structure to regularities in behavior which have shown to be valid in the real-world in real-time. It would be naive, however, to stick particular anatomical labels to the subcomponents of our models. They do provide, however, a perspective in which observations on properties of the neural substrate can be interpreted. In our example we will restrict ourselves to the further interpretation of DAC4.

Recently, the replay of neuronal firing patterns during sleep, in the same temporal order as during exploration, has been described in the rat hippocampus [39], [38]. Although these results have been questioned [29], the existence of two separate learning systems in the hippocampal loop and in the neocortex is well established [43]. The role of these separate learning systems has been investigated in abstract terms [26], but these investigations are limited since they do not rely on a model of learning which can be evaluated in the context of a behavioral task. A synthetic approach, as

demonstrated in this paper, allows such an evaluation. However, the model presented here is not inspired by the anatomy, physiology or neuropsychology of hippocampus and cortex, but addresses the general problem of communication between different neural structures, in the context of behaviorally realistic tasks and well evaluated models of learning.

We established that a system that is able to immediately acquire complex sequences and that can learn general properties of these sequences can be defined obeying the principles of locality, using two separate networks. The functional properties of these two networks are *a priori* not compatible and *a priori* not implementable in one homogeneous neural network. We demonstrated that a system relying on replay could combine the above features in a functionally valid way; this method is consistent with the functional requirements imposed by the external world, as discussed in section 4.

An interesting implication of the principle of locality is the use of replay; we use two different neural structures that describe the same sensory input and that need to exchange information. This exchange of information must be performed by synaptic transmission in order to respect the principle of locality. A system that would accomplish this exchange using direct connections, without using replay, would have to define a "code" for this transmission of information. Defining such a code, however, is unnecessary since the information encoded would describe the same sensory reality for both systems. Hence, replay provides a less complex solution, in which an internal code is not necessary.

However, using replay has some important implications for a neural system; during this phase, the neural populations in which sequences are replayed cannot be used for processing other inputs. This implies that these populations need to work in two exclusive modes; an *interactive-open* mode that allows sensory categorization, generation of action, acquisition of sequences by the short-term memory, and expression of sequences by the long-term memory, and a *passive-closed* mode where the neural populations have to be isolated from sensory inputs which would perturb the replay, in order to allow retention of the sequences.

It is obvious that an organism working with these two distinct modes is strongly weakened during the passive mode. One can wonder, however, why biological systems display circadian rhythms involving active and passive phases. If one assumes that a passive mode is necessary to a given biological process, like in our case the replay of activity patterns, then this mode can also be exploited by other processes, such as metabolic processes. The result of such a situation would be that these other mechanisms exploiting the passive mode, would in turn become dependent on it. Hence, it would be difficult to know *a posteriori* which process originally required a passive mode. Today, it is not clear whether replay is used in the brain [29], whether the corresponding passive mode is sleep, and whether sleep is necessary to other metabolic processes, like cell regeneration [22].

The modeling series presented in this paper is by no means complete

and is still under active development. It does illustrate, however, that a synthetic approach can provide insights in possible principles of neuronal organization and place them in relation to the overall behaving system, assuming that a number of conceptual and methodological considerations are met. It can provide a compensation for the more reductionistic methods followed in neuroscience with its implications of knowledge fragmentation. The validity of our own trajectory through the space of possible models needs to be scrutinized continuously and as such constitutes only an example of this approach. We do feel, however, that the problem of knowledge fragmentation does deserve the full attention of the field.

6 Appendix: Detailed specifications

6.1 The short-term memory

DAC4 is based on an earlier model of sequence learning which uses local learning rules [7]. Following this model the temporal context of short-term memory is represented by two neural populations, *State* and *Context*, which are recurrently connected. The integrated input, c_i, of unit i in *Context* at time $t+1$ depends on its input at t and on the input received from unit i in *State*, s'_i:

$$c_i(t+1) = (1-\alpha)\, c_i(t) + \alpha\, s'_i(t) \tag{4}$$

where α is a constant. The activity, c'_i, of unit i in *Context*, is a function f of its integrated input: $c'_i(t) = f(c_i(t))$, where f is a sigmoidal function. Next to inputs from *Context* units in *State* also receive external input from the CS population. The total input, s_i, of unit i in *State* is defined as:

$$s_i(t) = \sum_{j=1}^{N^C} w^C_{ij}\; c_j(t) + \sum_{k=1}^{N^{CS}} w^{CS}_{ik}\; CS_k(t) \tag{5}$$

where w^C and w^{CS} are random fixed synaptic weights. N^C and N^{CS} denote the size of populations *Context* and CS respectively, and CS_k is the activity of cell k in CS. The activity s'_i of unit i in *State* is a sigmoidal function f of its total input: $s'_i(t) = f(s_i(t))$.

The short-term memory uses a third population called *Segments*. Each of its units stores a sensorimotor couple in its synaptic connections with populations CS and UR, while its receptive field is a pattern of activity in *State*. The activity $g'_i(t)$, of unit i in *Segments*, is a Gaussian function of the Euclidean distance $g_i(t)$ between the actual pattern of activity in *State*,

$s'(t)$, and its synaptic weights:

$$g_i(t) = \left(\sum_{k=1}^{N^S} (w_{ik}^S - s'_k(t))^2 \right)^{1/2} \tag{6}$$

where w_i^S represents the vector of synaptic weights from *State* to unit i in *Segments* and N^S is the number of units in *State*, and:

$$g'_i(t) = \exp\left(-(g_i(t)/\sigma_i)^2\right) \tag{7}$$

where σ_i is the width of the Gaussian response of unit i. After updating the activities in *Segments* a winner take all competition selects the unit, k, with the highest activity:

$$g'_k(t) = \max_{i \in Segments} g'_i(t) \tag{8}$$

This mechanism involves non-local information. However, competition within a neural population can be locally implemented using lateral inhibition [21] and cannot be seen as a violation of our principle of locality.

The winning unit k in *Segments* updates its synaptic connections with *State*, CS and UR. This update is one-trial learning: the pattern of activity in *State* is immediately associated to the current sensorimotor couple through unit k. The new weight vector from *State* to k is: $w_k^S(t+1) = s'(t)$. In addition, the weight vectors w_k^{G-CS} and w_k^{G-UR} between the winning unit k in *Segments* and the CS and UR populations are modified according to:

$$\forall i \in 1...N^{CS}, w_{ki}^{G-CS}(t) = CS_i(t) \tag{9}$$

$$\forall i \in 1...N^{UR}, w_{ki}^{G-UR}(t) = UR_i(t) \tag{10}$$

where UR_i denotes cell i in UR. This one-step learning implies that the selected unit will loose a possible previous association. However, it is necessary to control how forgetting takes place in the short-term memory, because recently learned association need to be retained sufficiently long to allow retention by the long-term memory. The parameter σ_i in equation (7) controls the specificity of the response of unit i. A large value of σ_i means that unit i will respond to a wide range of stimuli. Modulation of this parameter allows the control of forgetting in the short-term memory. At each time step, σ_i is increased for all the units of *Segments*:

$$\sigma_i(t+1) = \lambda\, \sigma_i(t) \tag{11}$$

where $\lambda > 1$ is an increase rate. In addition, for the winner unit k, this width is reinitialized: $\sigma_k(t) = 1$. The loss of specificity (equation 11) ensures that units that have not been selected for a long time will have a higher probability to be selected in the future. In contrast, the probability that a recently selected unit will be selected again is low. This prevents disordered recruitment of the units in *Segments*.

6.2 Replay

Short-term memory patterns are retained in long-term memory through replay. During replay units in *Segments* are updated according to equations (7) and (8). The resultant winning unit in *Segments* will activate a new CS prototype. This will, in turn, lead to new activity in *State* and *Context*, and allows the next *Segments* unit to be selected (chaining). In this case the rate λ, and the parameter σ_i of each cell, remain at 1. In addition during replay units in *Segments* are able to activate units in UR.

6.3 The long-term memory

The long-term memory is implemented by a recurrent network, made of two populations, $STATE$ and $CONTEXT$. As in the case of the short-term memory, units of $STATE$ send one-to-one projections to the units of $CONTEXT$. They receive inputs from the CS and $CONTEXT$ populations. The connections from $CONTEXT$ to $STATE$ are updated as well as the connections from CS to $STATE$. The learning rule used is derived from the so called Self-Organizing Map algorithm [20]. In this case, however, it is applied to a recurrent network. This provides an adaptive representation of context [55].

The populations $STATE$ and $CONTEXT$ are two-dimensional. This allows to define the distance between two units. The internal activity, S_i, of unit i in $STATE$ depends on both the activity vectors $CS(t)$ in CS, and $C(t)$ in $CONTEXT$:

$$S_i(t) = \exp\left(-(a||W_i^{CS} - CS(t)|| + b||W_i^C - C(t)||)^2\right) \tag{12}$$

where W_i^{CS} is a vector representing the weights from CS to unit i of $STATE$, W_i^C the weights from $CONTEXT$ to i, a and b are real numbers, and $||...||$ denotes the Euclidean norm. The cell l of $STATE$ that has the highest activity is then selected:

$$S_l(t) = \max_{i \in STATE} S_i(t) \tag{13}$$

During replay, each unit i of $STATE$ has its synaptic weights updated, according to:

$$W_i^{CS}(t+1) = W_i^{CS}(t) + \gamma g_{il}(CS(t) - W_i^{CS}(t)) \tag{14}$$

$$W_i^C(t+1) = W_i^C(t) + \delta g_{il}(C'(t) - W_i^C(t)) \tag{15}$$

where γ and δ are learning rates, and g_{il} is a Gaussian function of the distance between units i and l. In addition, during replay, the current motor pattern is stored in the weights W^{UR} between the winner unit l in $STATE$ and the UR population:

$$W_l^{UR}(t) = UR(t) \tag{16}$$

During exploration, any cell of $STATE$ can activate the UR population if its activity is above a threshold. However, the learning rules defined in equations (14) and (15) change the receptive fields of the winning unit, but also of its neighbors. Therefore it is necessary to prevent units in $STATE$ from activating the UR population if their receptive field does not correspond to a learned sequence; the receptive field may have been modified more recently (equations 14, 15) than the output connections (equation 16). In this case, we limit the output activity of the cell so that it cannot excite the motor units, using a term $e_i(t)$ called excitability of unit i. Thus, the output activity of unit i of $STATE$ is defined as:

$$S_i'(t) = S_i(t)\ e_i(t) \tag{17}$$

If the output activity S_l' of the winner unit l is above a threshold, then it propagates its activity to the UR cells, inducing a motor action. During replay, the excitability e_l of the winner unit l is updated: If the replayed event corresponds to a non-default action, then this excitability is reset to one, $e_l = 1$, otherwise it decays, $e_l = (1 - \epsilon)\ e_l$, with $(0 < \epsilon < 1)$.

The activities of the cells in $CONTEXT$ depend on the nature of the

current action. If a non-default action is generated, then:

$$C_i(t+1) = (1-\beta)\, C_i(t) + \beta\, S_i(t) \tag{18}$$

where β is a constant. In the case of a default action the activity in $CONTEXT$ decreases, independently of the activity in $STATE$:

$$C_i(t+1) = (1-\beta)\, C_i(t) \tag{19}$$

This latter mechanism ensures that the pattern of activity in $CONTEXT$ continuously changes between two non-default actions. Therefore, units in $STATE$, that trigger non-default actions, can learn to respond after a certain time interval following the last non-default action triggered.

References

[1] Andrews JS. Possible confounding influence of strain, age and gender on cognitive performance in rats. Brain Res Cogn Brain Res. 1996 Jun; 3(3-4):251-67.

[2] Bear M, Malenka R. Synaptic plasticity: Ltp and ltd. Current Opinion in Neurobiology. 1994; 4:389-399.

[3] Bengio Y, Simard P, Frasconi P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. IEEE Transactions on Neural Networks. 1994; 5(2):157-166.

[4] Changeux J. Variation and selection in neural function. Trends Neurosci. 1997 Jul; 20(7):291-3.

[5] Chen C, Tonegawa S. Molecular genetic analysis of synaptic plasticity, activity dependent neural development, learning, and memory in the mammalian brain. Annu Rev Neurosci. 1997; 20:157-84.

[6] Dominey PF. Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning. Biological Cybernetics. 1995; 73:265-273.

[7] Dominey PF, Arbib M, Joseph JP. A model of Corticostriatal Plasticity for Learning Oculomotor Associations and Sequences. Journal of Cognitive Neuroscience. 1995; 7(3):311-336.

[8] Elman JL. Finding Structure in Time. Cognitive Science. 1990; 14:179-211.

[9] Elman JL, Zipser D. Discovering the hidden structure of speech. Journal of the Acoustical Society of America. 1988; 83:1615-1626.

[10] Galbicka G, Smurthwaite S, Riggs R, Tang L. (1997). Daily rhythms in a complex operant: targeted percentile shaping of run lengths in rats. Physiol Behav. 1997 Nov; 62(5):1165-9.

[11] Gallistel CR. The Organization of Learning. Cambridge, Ma.: MIT Press, 1990.

[12] Gardner H. The Mind's New Science: A history of the cognitive revolution. New York: Basic Books, 1987.

[13] Goldstein L, Smith K. Bugworld: A distributed environment for the study of multi-agent learning algorithms. Tech. Report, Department of Computer Science, UCSC, 1991.

[14] Greenough WT, Withers GS, Anderson BJ. Experience-dependent synaptogenesis as a plausible memory mechanism Learning and Memory: The behavioral and biological substrates. in Gormezano I and Wasserman EA eds. Hillsdale, NJ: Erlbaum, 1992; 7-24.

[15] Hodgkin A, Huxley A. A quantitative decsription of membrane current and its application to conduction and excitation in nerve. Journal of Physiology 1952; 127:500-544.

[16] Hull C. Principles of Behavior. New York: Appleton-Century-Crofts, 1943.

[17] Kamin L. "Attention-like" processes in classical conditioning. In Jones M, ed. Miami Symposium on the Prediction of Behavior: Aversive stimulation. Miami: University of Miami Press, 1967; 9-31.

[18] Kamin L. Predictability, surprise, attention, and conditioning. In Campbell, B. and Church, R., eds. Punishment and aversive behavior. New York: Appleton-Century-Crofts, 1969; 276-296.

[19] Kangas J. On the Analysis of Pattern Sequences by Self-Organizing Maps Dr. Tech. Thesis, Helsinki University of Technology, Laboratory of Computer and Information Science, 1994.

[20] Kohonen T. Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics, 1982; 43:59-69.

[21] Kohonen T. Self-Organized Maps. Springer-Verlag, 1997.

[22] Landis CA, Whitney JD. Effects of 72 hours sleep deprivation on wound healing in the rat. Res Nurs Health. 1997 Jun;20(3):259-67.

[23] Mackintosh NJ. The Psychology of Animal Learning. New York: Academic Press, 1974.

[24] Macphail E. The Neuroscience of Animal Intelligence: From the Seahare to the Seahorse. New York: Columbia University Press, 1993.

[25] Massaro D. Some criticism of connectionist models of human performance. Journal of Memory and Language. 1988; 27:213-234.

[26] McClelland J, McNaughton B, O'Reilly R. Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionnist Models of Learning and Memory. Psychol Rev. 1995 Jul; 102(3):419-57.

[27] Miller RR, Barnet RC, Grahame NJ. Assessment of the Rescorla-Wagner model. Psychol Bull. 1995 May; 117(3):363-86.

[28] Moore ME. Gedanken-experiments on sequential machines. In Shannon CE and McCarthy J eds. Automata Studies. Princeton: Princeton University Press, 1956; 129-153.

[29] Moore GP, Rosenberg JR. (1996). "Replay" of Hippocampal "Memories". Science. 1996 nov 15; 274(5290):1216-7.

[30] Mondada F, Verschure PFMJ. Modeling system-environment interaction: The complementary roles of simulations and real world artifacts. Proceedings of the Second European Conference on Artificial Life. Cambridge, Ma.: MIT press, 1993; 808-817.

[31] Newell A. Unified Theories of Cognition. Cambridge, Ma.: Harvard University Press, 1990.

[32] Oja E. A simplified neuron model as a principal component analyzer. Journal of Mathematical Biology. 1982; 15:267-273.

[33] Pavlov IP. Conditioned Reflexes. Oxford University Press, 1927.

[34] Pearlmutter BA. Gradient calculation for dynamic recurrent neural networks: a survey. IEEE Trans. Neural Networks. 1995; 6:1212-1228.

[35] Piaget J. Biology and Knowledge: An essay on the relations between organic regulations and cognitive processes. Chicago: University of Chicago Press, 1971.

[36] Rescorla RA, Wagner AR. A Theory of Pavlovian Conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. in A.H.Black and W.F.Prokasy eds. Classical Conditioning 2, current theory and research. New York: Appleton-Century-Crofts, 1972; 64-99.

[37] Romanes GJ. Animal Intelligence. New York: Appleton, 1888.

[38] Shen J, Kudrimoti HS, McNaughton BL. Reactivation of neuronal ensembles in hippocampal dentate gyrus during sleep after spatial experience. J. Sleep Research 1998; 7 suppl 1:6-16.

[39] Skaggs WE, McNaughton BL. Replay of Neuronal Firing Sequences in Rat Hippocampus During Sleep Following Spatial Experience. Science. 1996 Mar 29; 271(5257):1870-3.

[40] Smith E, Jonides J. Neuroimaging analyses of human working memory. Proc Natl Acad Sci USA. 1998 Sep 29 ;95(20):12061-8.

[41] Strohman R. The coming Kuhnian revolution in biology. Nature Biotechnology. 1997; 15:194-200.

[42] Squire LR. Memory and Brain. Oxford university Press, 1987.

[43] Squire LR, Knowlton B, Musen G. The Structure and Organization of Memory. Annu Rev Psychol. 1993; 44:453-95.

[44] Thorndike EL. Animal Intelligence: an experimental study of the associative processes in animals. The Psychological Review Series of Monograph Supplements. New York: Macmillan, 1898; 8.

[45] Verschure PFMJ. Taking connectionism seriously: The vague promise of subsymbolism and an alternative. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, Bloomington, Indiana. Hillsdale, N.J.: Erlbaum, 1992; 653-658.

[46] Verschure PFMJ. The cognitive development of an autonomous behaving artifact: The self-organization of categorization, sequencing, and chunking. In Cruze H, Ritter H, Dean J eds. Proceedings of the Second Brazilian International Conference on Cognitive Science, 1993.

[47] Verschure PFMJ. Connectionist explanation: Taking positions in the mind-brain dilemma. In Dorffner G, ed. Neural Networks and a New Artificial Intelligence, pp. 133-188. London: Thompson, 1997. First presented at ZiF workshop Mind and Brain, 1990.

[48] Verschure PFMJ, Coolen ACC. Adaptive fields: Distributed representations of classically conditioned associations. Network. 1991; 2:189-206.

[49] Verschure PFMJ, Kröse B, Pfeifer R. Distributed adaptive control: The self-organization of structured behavior. Robotics and Autonomous Systems. 1992; 9:181-196.

[50] Verschure PFMJ, Pfeifer R. Categorization, representations, and the dynamics of system-environment interaction: a case study in autonomous systems. In Meyer JA, Roitblat H, Wilson S eds. From Animals to Animats: Proceedings of the Second International Conference on Simulation of Adaptive behavior, Honolulu:Hawaii. Cambridge, Ma.: MIT press, 1992; 210-217.

[51] Verschure, P.F.M.J., Wray, J., Sporns, O., Tononi, G., and Edelman, G. Multilevel analysis of classical conditioning in a behaving real world artifact. Robotics and Autonomous Systems. 1995; 16:247-265.

[52] Verschure PFMJ. Xmorph: A software tool for the synthesis and analysis of neural systems. Tech Report Institute of Neuroinformatics, ETH-UZ, 1997.

[53] Verschure PFMJ, Voegtlin T. A bottom up approach towards the acquisition and expression of sequential representations applied to a behaving real-world device: Distributed Adaptive Control III. Neural Networks. 1998; 11:1531-1549

[54] Verschure PFMJ. Synthetic Epistemology: The acquisition, retention, and expression of knowledge in natural and synthetic systems. Proceedings of the IEEE World Congress on Computational Intelligence 1998; 147-153.

[55] Voegtlin T, Dominey PF. Contextual Self-Organizing Maps: An adaptive representation of context for sequence learning. Tech Report of the Institut des Sciences Cognitives, Lyon, France. 1998-01.

[56] Wang D, Yuwono B. Anticipation-Based Temporal Pattern Generation. IEEE Trans on Systems, Man, and Cybernetics. 1995; 25:615-628.

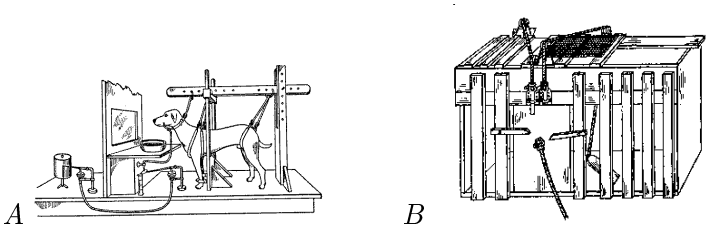


Figure 1: A: Experimental setup used by Pavlov. B: Thorndike's puzzle box.

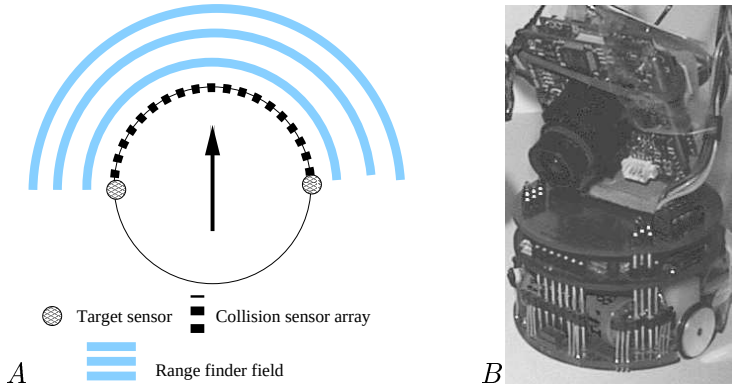


Figure 2: A:The soma of the simulated robot: Target sensors are placed on each side. The front side is covered with collision sensors and distal sensors. Arrow indicates the primary direction of motion. B: The Khepera robot. The CS is the image from the color CCD camera mounted on top. The US comes from light and IR sensors placed at the lower circumference of the base. The Khepera robot is circular with a diameter of 3 cm and 8 cm high, including the camera.

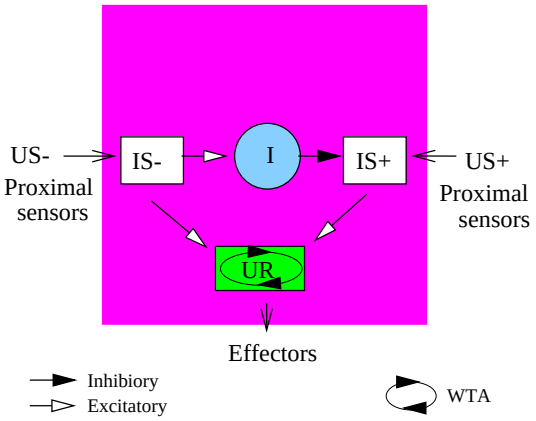


Figure 3: The reactive control structure. Collision sensors and target sensors (US) modify the internal state (IS). The IS populations trigger reactive motor actions. An inhibitory unit defines the priority between the two IS populations.

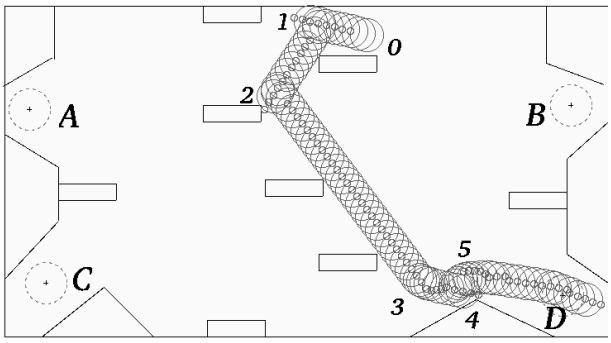


Figure 4: A trajectory of the soma generated by a reactive control structure. A collision triggers an avoidance reaction (US^-) (1,2,4). Targets in A,B,C,D emit a signal that can be detected by the sensors (appetitive unconditioned stimulus, US^+). These can trigger approach actions (3,5).

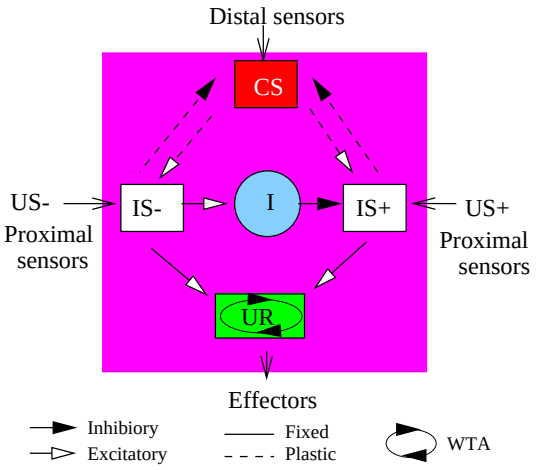
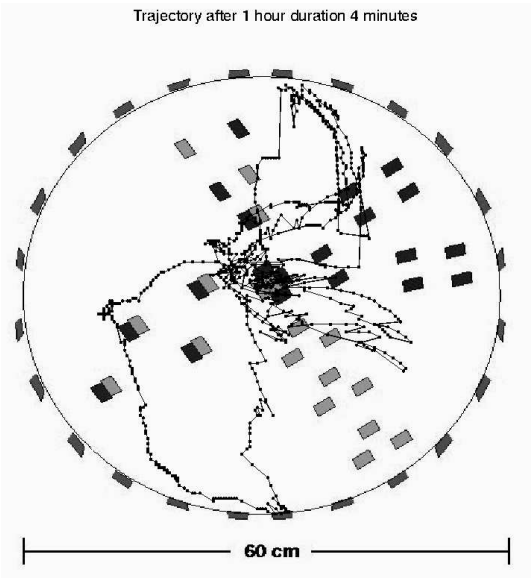


Figure 5: The adaptive control structure. A recurrent loop with inhibitory feedback connections allows to learn how to categorize the CS events. The CS modifies the internal states (IS), triggering conditioned reflexes.

A



B

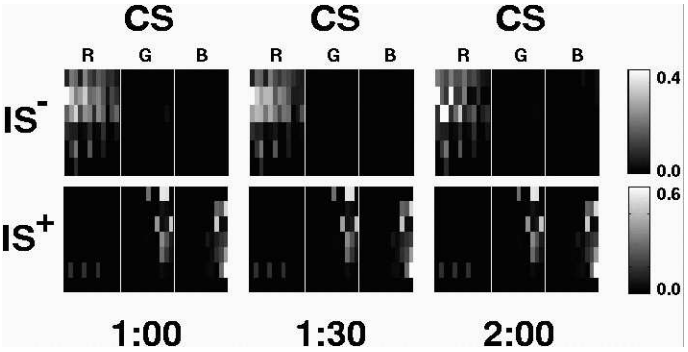


Figure 6: A: An environment used for the Khepera robot. Blue and green color patches are dispersed on the floor (light and dark gray respectively on the figure), and the environment is delimited by a circular wall on which red patches are attached. Red was correlated with collisions (US-) while green and blue were correlated with the presence of a light source (targets-US+) placed over the middle of the environment. A trajectory of the robot is plotted, which lasted 4 minutes and was recorded after one hour of exploration. B: Synaptic connections between the color responsive CS cells and the IS populations. Each column represents the set of cells of CS responsive to a specific color, red, green, and blue respectively. The first row of matrices represents the connections between the cells of the $IS-$ population and the color cells. The lower row displays the strength of the connections between $IS+$ and the color sensitive cells. The first display shows the connectivity pattern after 1 hour of learning the subsequent displays relate to the connectivity after 1.5 and 2 hours. Each row in a sub-matrix can be interpreted as the receptive field of the IS neurons. The top row in each sub-matrix corresponds with the sensor placed at 90^o of the center of the robot. Each following sensor is placed at -30^o from the previous one. The last two rows correspond with the sensors placed at the back of the robot.

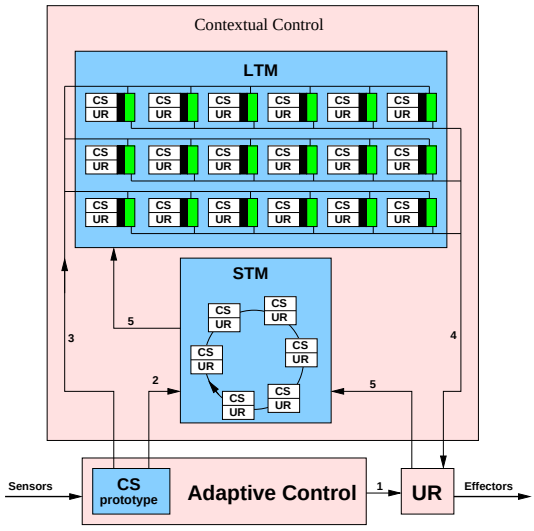


Figure 7: The contextual control structure of DAC3. 1: The UR population receives inputs from the IS population of the adaptive control structure. 2: If a non-default action occurs, the CS prototype and the UR activity are stored as a segment in the short-term memory. 3: The current CS prototype is matched against prototypes of the segments in the long-term memory. 4: If a CS prototype in the long-term memory matches the current CS prototype, then then contextual control structure induces a motor action. 5: If a sequence is selected, the segments in the short-term memory are stored in long-term memory.

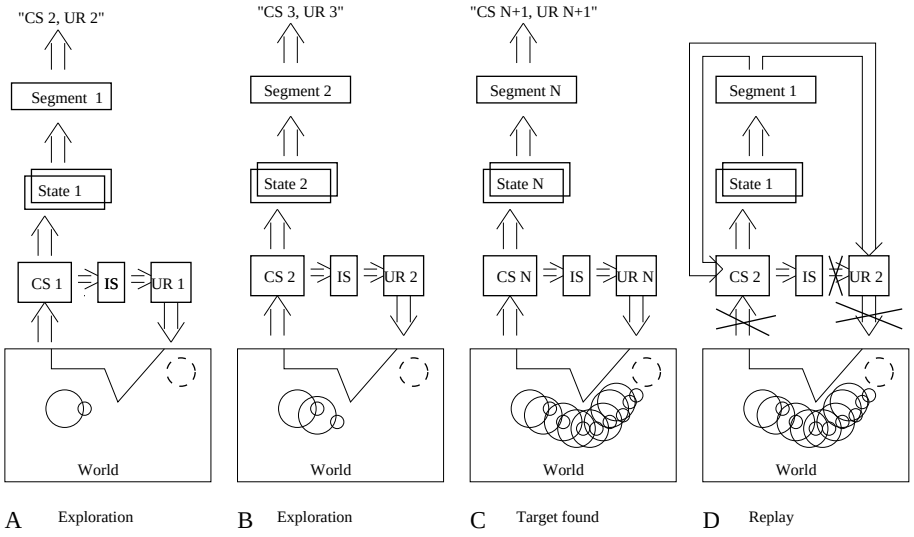


Figure 8: The short-term memory of DAC4. The *CS*, *IS*, *UR*, *State*, *Context*, and *Segments* populations are represented. The recurrent connections between *State* and *Context* are symbolized by the double frame labeled "State". A cell of the *Segments* population associates a pattern of a recurrent network with the sensorimotor events of the next time step. A,B: Exploration. In A, the state of the recurrent network that depends on the current context (CS1, and before), is associated with the stimulus and the motor action of the current time step (CS2, UR2). C: A target is found, this will trigger replay to allow retention in LTM. D: Replay was initiated by *Segments* unit 1. During the replay, the selected cell of *Segments* (unit 1) activates the CS population with the associated CS events (CS2), in order to generate the next pattern in the recurrent network (State2). The *UR* population is also activated in order to allow its acquisition by the long-term memory. Motor actions are inhibited during replay.

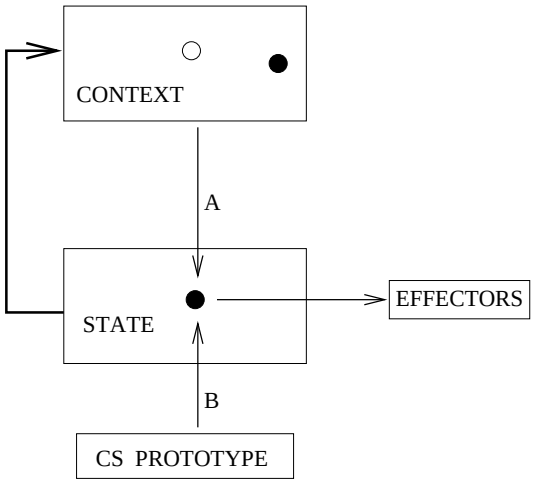


Figure 9: Long-term memory of DAC4: A cell of $STATE$ has a double receptive field, one part corresponding to the temporal context (A) and the other part to the stimulus (B). The context layer $CONTEXT$ is in turn activated by $STATE$. A competition in $STATE$ selects the cell responding to both the input (CS prototype) and the context represented in $CONTEXT$.

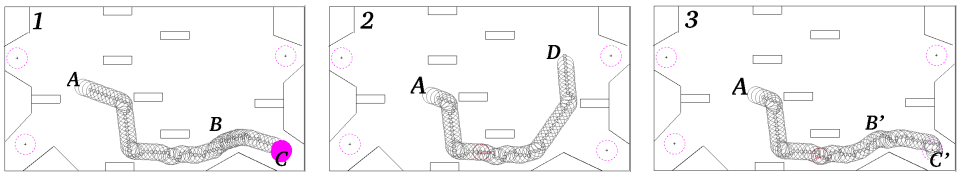


Figure 10: A sequence successfully learned by the contextual control of DAC4. 1: (*Stimulation period*) The soma starts from (A). The target in the right lower corner emits a signal which attracts the soma at location (B), until the target is found (C). 2: (*Recall period before learning*) The signal coming from the targets has been removed. The soma of DAC4, started from (A) does not find the target (D). 3: (*Recall period after learning*) The long-term memory of DAC4 expresses the sequence. The soma starts from (A). In (B') the contextual control structure engages approach actions learned in (1), and the target is found in (C').