Adaptive **Behavior**

Neuromodulation of Reactive Sensorimotor Mappings as a Short-Term Memory Mechanism in Delayed Response Tasks

Tom Ziemke, Mikael Thieme
Department of Computer Science, University of Skövde

This article addresses the relation between memory, representation, and adaptive behavior. More specifically, it demonstrates and discusses the use of synaptic plasticity, realized through neuromodulation of sensorimotor mappings, as a short-term memory mechanism in delayed response tasks. A number of experiments with extended sequential cascaded networks, that is, higher-order recurrent neural nets, controlling simple robotic agents in six different delayed response tasks are presented. The focus of the analysis is on how short-term memory is realized in such control networks through the dynamic modulation of sensorimotor mappings (rather than through feedback of neuronal activation, as in conventional recurrent nets), and how these internal dynamics interact with environmental/behavioral dynamics. In particular, it is demonstrated in the analysis of the last experimental scenario how this type of network can make very selective use of feedback/memory, while as far as possible limiting itself to the use of reactive sensorimotor mechanisms and occasional switches between them.

Keywords delayed response tasks · short-term memory · neuromodulation · synaptic plasticity · higher-order recurrent neural nets · nonrepresentational memory

1 Introduction

Delayed response tasks, which are common in experimental psychology research, have recently also received much attention in the adaptive behavior and artificial neural network (ANN) research community (e.g. Ulbricht, 1996; Jakobi, 1997; Rylatt & Czarnecki, 1998, 2000; Linåker & Jacobsson, 2001, 2002; Bergfeldt & Linåker, 2002; Thieme & Ziemke, 2002). A simple example of a delayed response task is the one illustrated in Figure 1, which has also been referred to as the "road sign problem" (Rylatt & Czarnecki, 2000). Here an agent starts off at the bottom/root of a simple T-shaped maze, encounters an instruction stimulus (e.g., a light) while moving along a corridor, and after some further delay, reaches a T-junction at which the correct turning direction depends on where, that is, on which side, the stimulus was encountered (typically the agent is expected to turn toward the same side).

Delayed response tasks are a standard way of investigating short-term memory (STM). The agent is typically assumed to "remember" in some way the necessary information about the stimulus (in the above case the side on which it appeared) during the delay period. None of the synthetic studies mentioned above, all of them using simple simulated or physical robotic agents, was aimed at providing a specific model of how animals solve delayed response tasks,

Correspondence to: T. Ziemke, Department of Computer Science, University of Skövde, Box 408, 54128 Skövde, Sweden.
E-mail: tom@ida.his.se; *Tel.:* +46-500-438330; *Fax:* +46-500-438399

(2002), Vol 10(3–4): 185–199.
[1059–7123 (200210) 10:3–4;185–199; 033698]

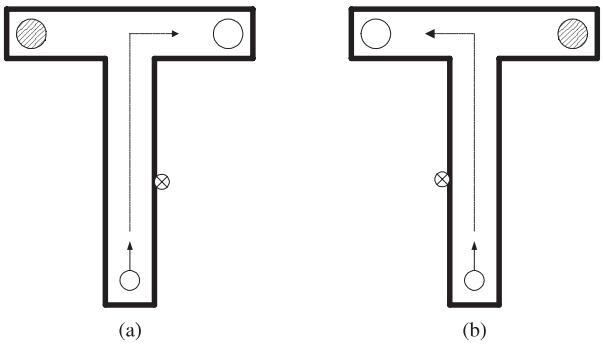

Figure 1 The two situations in the T-maze environments of the road sign problem. Adapted from Ulbricht (1996). The empty circles indicate the goals, whereas the striped circles indicate areas the agent should not enter.

and neither is the work presented here. Nevertheless, this type of study can shed light on some of the issues involved, such as the nature and the role of the assumed STM mechanism. For example, from an observer's point of view it is relatively easy to attribute some kind of "representation" to an animal or robot exhibiting the correct behavior, but the detailed analysis of synthetic studies can help to illuminate the actual mechanisms underlying that behavior.

Most of the above synthetic studies used standard recurrent ANNs in which certain neuron activation values are fed back and used as extra inputs to some of the neurons in a later time step (typically the next one). In this type of network the synaptic connection weights are usually considered *long-term* memory since they are changed only by the training process, whereas the feedback of activation values, which can change from moment to moment, is commonly considered to constitute *short-term* memory. Accordingly, most of these studies conceive of STM, as is common in much ANN research, as realized through the feedback of more or less stable patterns of neuronal activation. In delayed response tasks such patterns can be triggered or "created" when the stimulus is encountered and sustained during the delay period at the end of which they guide the robot's behavior as some kind of "stand-in" for the original stimulus.

This focus on neural activity as the basis of STM is largely consistent with neuroscientific findings. As Durstewitz, Seamans, and Sejnowski (2000) pointed out in a recent review of neurocomputational models of working memory, memory-related delay activity has been observed in several brain areas, in particular the prefrontal cortex (PFC), "the brain structure most closely linked to working memory" (cf., for example, Fuster, 1973; Goldman-Rakic, 1987; Guigon & Burnod, 1995). This has been found in both single-neuron recordings in nonhuman primates and in human brain imaging studies (cf., for example, Baddeley, 1998; Wang, 2001). Durstewitz et al. (2000) further elaborate:

> PFC neurons show elevated persistent activity during delayed reaction tasks, when information derived from a briefly presented cue must be held in memory during a delay period to guide a forthcoming response … Thus, this type of short-term memory relies on the maintenance of elevated firing rates in specific subpopulations of neurons rather than on synaptic plasticity, which might underlie long-term memory. (Durstewitz et al., 2000, p. 1184)

Although synaptic plasticity is largely neglected in most current neuroscientific and ANN/robotic models of STM we believe that there are good reasons for paying closer attention to its possible role. Firstly, as Durstewitz et al. (2000) point out, "different cellular and network mechanisms…are not mutually exclusive." This means, even if much is known about the role of persistent neural activity in STM, this does not in any way rule out a possible role for synaptic plasticity as at least a complementary mechanism. Durstewitz et al. (2000), for example, speculate that "through mechanisms for synaptic plasticity, more permanent representations…might be formed that enhance robustness of sustained activity and enable fast processing at lower, metabolically economical firing rates." Secondly, as Durstewitz et al. (2000) also point out, the role of neuromodulators such as dopamine is simply not understood yet, although it has been observed that, for example, dopaminergic activity, which can effect synaptic currents, increases during working memory tasks (e.g., Schultz, Apicella, & Jungberg, 1993; Watanabe, 1996). Thirdly, currently available techniques for observing brain activity, such as single-neuron recordings and brain imaging techniques, focus on the measurement of neuronal activity, whereas synaptic changes are more difficult to monitor. Hence, it is hardly surprising that the role that persistent neural activity plays in working memory is better studied and documented than the possible role of synaptic plasticity. In synthetic studies, however, both are equally easy or difficult to model, at least at an abstract level.

Hence, we believe that synthetic studies of the role of synaptic plasticity in STM can be interesting from both a scientific and an engineering perspective. The former because they might contribute to neuro- and/or cognitive-scientific theories and models of the corresponding biological mechanisms, and the latter because they might provide new ideas for implementing STM, for example, in robots.

This article therefore aims to illustrate an artificial neural STM mechanism that is not based on the feedback or sustenance of neuronal activation, but on the dynamic modulation of synaptic connection weights in so-called higher-order recurrent ANNs. The workings of this mechanism are illustrated in experiments with a simple simulated robot facing the above delayed response task and more complex variations of it (cf. also Thieme, 2002; Thieme & Ziemke, 2002). As in most adaptive behavior research, the ANNs used here are only very rough abstractions of actual or possible biological mechanisms. Hence, this article certainly does not directly contribute to neuroscientific theories or models, although it will, for example, raise the question if STM necessarily needs to be representational. The article aims to illuminate further, at a more abstract level, the potential role that neuromodulation of synaptic weights can play in STM, and thus aims to contribute potentially to the development of theories and models of the underlying mechanisms in biological systems.

The rest of this article is structured as follows: The next section provides some background on different types of feedback in recurrent ANN architectures and the way they can be used as STM mechanisms. Furthermore, the relation to other work on dynamic short-term adaptation/modulation of sensorimotor mechanisms is addressed. Section 3 describes the experiments, and Section 4 analyzes the most relevant results with a focus on the use of neuromodulation of connection weights, resulting in adaptive sensorimotor mappings, as a STM mechanism in these tasks. Section 5, summarizes the article and presents some conclusions.

2 Background

2.1 Feedback in Recurrent Neural Networks

Recurrent ANNs are commonly used in autonomous agents and adaptive behavior research because they offer a uniform, low-level mechanism allowing the integration of a more or less direct mapping from sensory input to motor output with an implicit handling of both long- and short-term memory. The majority of recurrent neural architectures used in this type of research make use of *first-order feedback*. As mentioned above, this means certain neuron activation values are fed back into the network, typically in the next time step. Meeden (1996), for example, trained networks of this type to control a robot to switch periodically between approaching and turning away from a light source. The robot managed to solve the task even when not given any input providing information about its current goal/subtask. Using its internal feedback connections to remember its current goal, it was able to deal with a situation of perceptual aliasing, and thus could respond differently to identical light sensor inputs in different contexts, that is, approach or turn away depending on what the current goal was. Similarly, Ziemke (1999) showed how first-order recurrent ANNs controlling a robot trained to avoid objects in one part of environment, but to hit identical objects in another part, solved the task by feeding back a number of different neural activation patterns to remember which part of the environment they were currently in.

In the case of *higher-order feedback*, on the other hand, it is typically connection weights (and/or bias weights) that are modulated dynamically. Examples of architectures utilizing this type of feedback are Pollack's (1987, 1991) sequential cascaded network (SCN) and our variation, the extended SCN (ESCN; Ziemke, 1999, 2000) which is illustrated in Figure 2. Both of these architectures can be described as consisting of a *function network* that maps inputs to output and state units using a single layer of connection weights (the function network weights), and a *context network* that takes as inputs the state unit activation values and produces as output the next time step's function network weights.

More specifically, given an input vector $i_j(t)$, $j = 1\ldots n$, state unit vector $s(t)$ and output vector $o(t)$ are usually calculated as follows by the *function network*:[1]

$$o(t) = f(W_o(t) \cdot (i_1(t),\ldots,i_n(t), 1))$$
$$s(t) = f(W_s(t) \cdot (i_1(t),\ldots,i_n(t), 1))$$

Where f is the logistic activation function, and $W_o(t)$ and $W_s(t)$, together referred to as *function network weights*, are two-dimensional connection weight matrices dynamically computed in every time step by

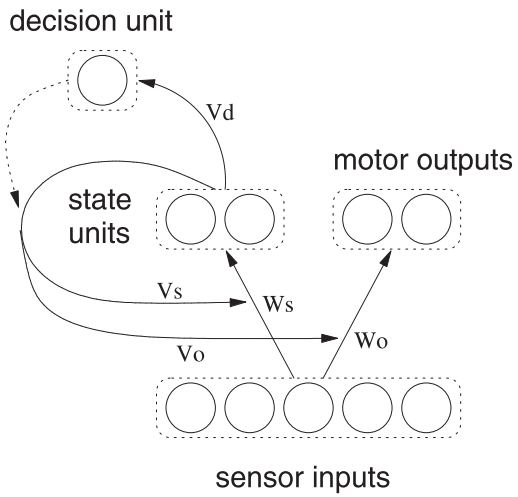

Figure 2 Extended sequential cascaded network (ESCN) as a robot controller. Solid arrows indicate fully connected connection weight matrices. The dotted arrow indicates the selectivity of feedback, depending on the activation of the decision unit.

the *context network*. This works slightly differently in the ESCN than in the SCN. The difference is that in the SCN the context network is used in every time step, whereas its use is conditional upon the activation of an additional *decision unit* in the ESCN (cf. Figure 2). The idea behind this extension is that the robot should be able to decide actively and selectively when to change its sensorimotor mapping, instead of (re-) setting the function network weights in each and every time step. More specifically, given the state vector $s_k(t)$, $k = 1 \ldots m$, the decision unit activation $d(t)$ and the function network weight matrices $W_o(t)$ and $W_s(t)$, are dynamically computed in every time step t as follows by the ESCN's *context network*:

$$d(t) = f\,(V_d \cdot (s_1(t),\ldots,s_m(t), 1))$$

$$\text{if } d(t) \geq 0.5 \text{ then } W_o(t+1) = V_o \cdot (s_1(t),\ldots,s_m(t), 1)$$

$$\text{else } W_o(t+1) = W_o(t)$$

$$\text{if } d(t) \geq 0.5 \text{ then } W_s(t+1) = V_s \cdot (s_1(t),\ldots,s_m(t), 1)$$

$$\text{else } W_s(t+1) = W_s(t)$$

where f is the logistic activation function, V_d is a one-dimensional connection weight matrix, mapping current state to decision unit activation, and V_o and V_s are two-dimensional connection weight matrices mapping the current internal state $s(t)$ to the next time step's function network weights, if the decision unit is active, that is, has an activation value of at least 0.5.

Unlike in other areas, such as (formal) language recognition, there are only relatively few cases where higher-order networks have been used in autonomous agents and adaptive behavior research. One reason might be that first- and higher-order networks are computationally equivalent (Siegelmann & Sontag, 1995; Siegelmann, 1998), that is, every task solved by a higher-order network could also, at least in theory, be solved by some first-order net, and vice versa. Computational equivalence of network architectures in theory, however, does not say much about their suitability to solve particular tasks in practice, or to serve as a model of the biological neural mechanisms underlying memory. Given that first-order recurrent networks are already relatively well understood, the rest of this article focuses on higher-order networks and the way they can realize STM in delayed response tasks.

2.2 Related Work

To our knowledge, the work presented here, along with our earlier related work (e.g. Ziemke, 1996, 1999, 2000), is unique in its use of higher-order recurrent networks and in the type of analysis presented here. However, at least four groups/types of related work on evolved synaptic plasticity and "fast," that is, moment-to-moment, adaptation of sensorimotor mappings can be identified. These will be addressed briefly in the following, in order of increasing relatedness.

Firstly, there is work on the evolution of *synaptic plasticity*, that is, evolved neurocontrollers that change their connection weights during their "lifetime" in interaction with the environment. Floreano and Mondada (1996), for example, evolved their neurocontrollers' use of different Hebbian-style rules for fast moment-to-moment "learning" (cf. also Nolfi & Floreano, 2000). The controlled robots exhibited stable behavior realized through the continuous change of connection weights in a dynamically stable and apparently coordinated fashion. Contrary to our work, however, this weight fluctuation is based on *local* learning rules, that is, each connection weight changes moment-

to-moment dependent on the activation of the pre- and post-synaptic neurons. This means there is no possibility of (or perhaps no need for) global modulation, as realized by the context network in our higher-order networks; furthermore there is no possibility of selective use of feedback and weight fluctuation as in our extended SCN.

Secondly, there is another recent approach to the use of *global modulation*. The work of Husbands, Smith, Jakobi and O'Shea (1998) on so-called *GasNets* is strongly inspired by the modulatory effects of diffusing gases in biological neural networks. Applying this to robot controllers, an evolutionary algorithm was used to construct control networks of gas-emitting and conventional neurons, each of which had a certain position in a two-dimensional plane. An abstract model of the temporal and spatial properties of gas diffusion was used, and the conventional neurons' transfer functions were effected by the current concentration of gas. This type of mechanism was used for training a camera-equipped robot on target discrimination (approach of triangles on the wall, avoidance of rectangles). In this case, the evolutionary process determined the two-dimensional network structure (including the visual morphology, that is, which camera pixels to use as input) and the settings of the synapses, whereas during the individuals' lifetime the conventional neurons/synapses could be modulated dynamically. Hence, GasNets can roughly be likened to our higher-order recurrent neural robot controllers as follows. In both cases there is one subnetwork embodying a sensorimotor mapping, and there is another global mechanism that dynamically modulates that sensorimotor mapping (through gaseous modulatory feedback and context-network-feedback respectively). However, Husband et al.'s work is clearly much more biologically inspired than ours. Moreover, there are significant differences in the way feedback is used, both in the mechanism of adaptation (gas diffusion vs. instantaneous adaptation) and in its target (transfer functions vs. connection weights).

Thirdly, another recent approach to the use of global modulation is the work of Ishiguro et al. (2000) on the use of *neuromodulators* in *dynamically rearranging neural networks*. This work has been inspired by Meyrand, Simmers, and Moulins' (1991) work on the lobster's stomatogastric nervous system, which suggests that to some degree "biological nervous systems are able to change dynamically their structure as well as their synaptic weights" (Ishiguro et al., 2000). Using evolutionary algorithms, controllers for a legged robot were constructed. As in the work of Husbands et al. (1998), the diffusion of neuromodulators was evolved together with the network architecture. Moreover, neurons were supplied with specific receptors, sensitive only to certain neuromodulators, which made the modulation more selective. Concerning the relation to our work, similar comments apply as in the case of Husbands et al. (1998). At an abstract level, there are functional similarities in the way a sensorimotor mechanism is combined with a second, modulatory mechanism. At the level of implementation details, however, there are numerous differences, most obviously in the level of biological detail integrated into the modulatory mechanism.

Finally, there is the work of Bergfeldt and Linåker (2002), which is most closely related to our work (and also comes from the same lab). Bergfeldt and Linåker constructed a two-level ANN architecture that consisted of two mechanisms that are roughly equivalent to context and function network in SCN and ESCN. The lower-level sensorimotor mapping consisted of a simple feed-forward network with input units and output units connected by one layer of weights. The higher level consists of an unsupervised competitive learning mechanism that abstracts from the sensory input to simple concepts like "corner" or "corridor," and a modulation network that maps the currently active concept to values that are added to the lower-level's motor biases. Using an evolutionary algorithm, systems of this type have been trained successfully on simple delayed response tasks. The differences and similarities to the work presented in this article are fairly obvious: Functionally and structurally their two-level architecture is very similar to SCN and ESCN, but the modulation is realized in a very different manner and only applied to motor biases rather than to all sensorimotor connection weights.

3 Experiments

It might be worth noting that the experiments documented here are part of a larger set of experiments with four different ANN architectures (Thieme, 2002; Thieme & Ziemke, 2002). This article focuses on the experiments with ESCN robot controllers. For details on the other experiments and for a more detailed

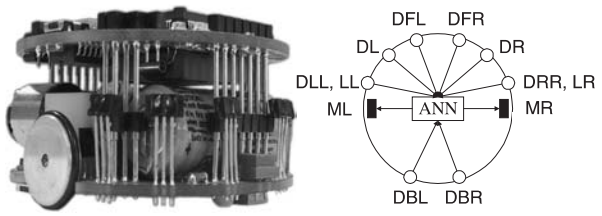

Figure 3 The Khepera robot (Mondada, Franzi, & Ienne, 1994) and its simulated counterpart. The labels beginning with D refer to the infrared distance/proximity sensors: back left (DBL), left-left (DLL), left (DL), front left (DFL), front right (DFR), right (DR), right-right (DRR), and back right (DBR). The figure also shows left and right wheel/motor (ML and MR), which are controlled independently, as well as the left and right ambient light sensors (LL and LR) used in the experiments.

quantitative evaluation the reader is referred to Thieme (2002).

3.1 Agent and Environment

The experiments discussed in this article have been carried out with a simulated Khepera robot (cf. Figure 3), using an extended version of Michel's (1996) Khepera simulator. The robot faced the six variations of the road sign problem illustrated in Figure 4. In S1, the original T-maze problem, the agent should turn to the side where the stimulus appeared, whereas in S2, a simple variation, it should turn to the other side. R1 and R2 are referred to as *repeated T-maze problems*. In R1, the agent should turn to the side(s) where the stimulus appeared, possibly twice. In R2, on the other hand, the first and second light have different meanings, that is, in the second case the agent should turn toward the other side. M1 and M2 are referred to as *multiple T-maze problems*, that is, here *both* stimuli come before the first T-junction. In M1, both turns should go toward the side where the respective stimulus appeared, whereas in M2, as in R2, the meaning of the second light is reversed, that is, the robot should turn to the other side.

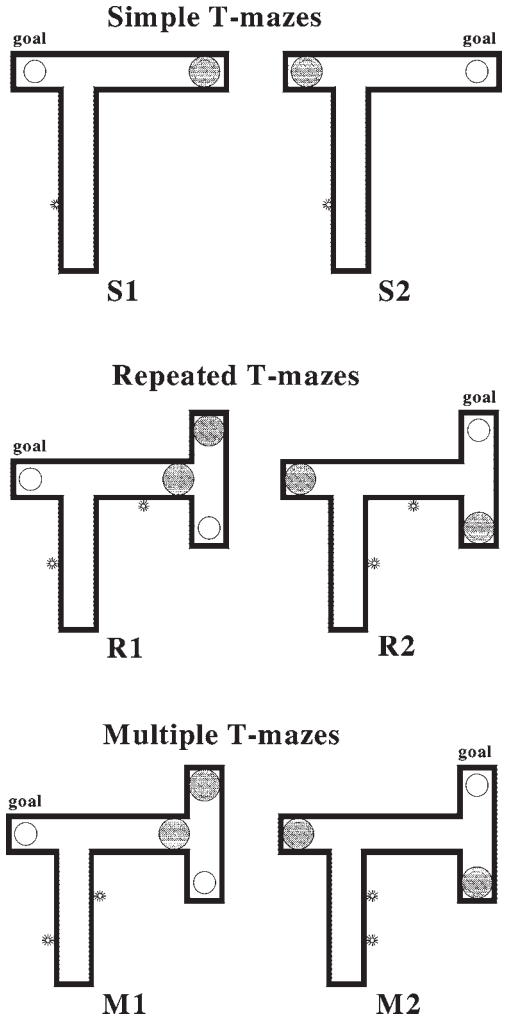

Figure 4 Example situations in six variations of the road sign problem (exact start position, orientation and light locations vary randomly, and goal locations vary accordingly). Empty circles indicate goal locations, whereas gray/striped circles indicate areas in which the agent "dies" immediately.

3.2 Control Networks and Training

Networks of the ESCN architecture, as illustrated in Figure 2 and discussed in Section 2.1, were used as robot controllers, receiving sensory input from the robot's eight infrared distance/proximity sensors and two light sensors (cf. Figure 3), and controlling its motors through two output units. As part of the simulator's functionality, random noise was added to simulated sensor and motor values as follows: ±10% to distance sensor values, ±5% to light sensor values, ±10% to motor speed amplitudes, and ±5% to the direction of motion resulting from the speed differences (Michel, 1996).

Networks were trained, for each of the six environments separately, by evolving the connection weights over 1,000 generations using a fairly standard evolutionary algorithm. Agents were selected based on their capacity to reliably reach the goal within a

limited number of time steps without touching the walls. In each fitness evaluation the agent was faced with different situations in the environment in a cyclic fashion (starting in a random situation) until it had solved a maximum of 20 situations or until it failed, either by entering a dead end, touching a wall, or not reaching the goal within 400 time steps. The fitness measure was simply the number of situations solved out of the maximum of 20 situations. It should be noted that each general situation, that is, each combination of possible light positions (left-right, left-left, etc.), could appear several times since, apart from random variations in distances between starting position, light position and T-junction, there are at most four general situations in each environment.

Each artificial genome consisted of a sequence of real values in the range [–10,10] that were mapped into context network weights in a one-to-one fashion. The input-output weights (in the function network), however, were not evolved but determined by propagating two initial context/state unit activations (evolved along with the context network weights) through the context network into initial settings for these weights. The algorithm used a rank selection mechanism that repeatedly picked out the best individual out of two individuals randomly chosen among the 20 fittest in the population, consisting of 100 individuals in total. Each selected individual reproduced asexually (i.e., without crossover), resulting in a single offspring by means of a mutation mechanism, in which each weight was offset, with a probability of 0.05, by a value in the range [–8,8] (but kept in the interval [–10,10]). An elitist mechanism was also used which preserved the best individual unmodified.

The starting condition for the agent and the placement of the light sources in each configuration were varied to rule out brittle solutions relying heavily on these aspects. The starting position was offset by a value in the range [–*r*, *r*], where *r* is the radius of the agent, along the *x*- and *y*-axis, respectively. The starting angle was chosen randomly between 70 and 110º, where an angle of 90º means that the agent faces straight toward the first junction. The placement of each light source was offset along the corridor (i.e., either along the *x*- or *y*-axis) by a value in the range [–*d*, *d*], where *d* is the diameter of the robot.

4 Results and Analysis

ESCN controllers evolved to reliably reach the goal in all environments. To illustrate how these networks work, we will go through successful representative solutions to all six T-mazes, covering some of them in detail, and focusing on the use of neuromodulation of sensorimotor mappings as an STM mechanism. For a

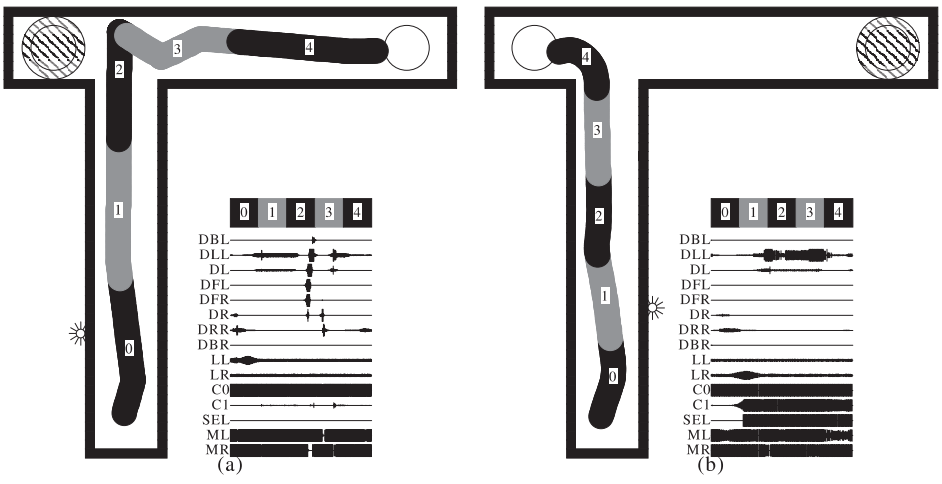

Figure 5 ESCN-controlled agents' behavior and activation values in the second simple T-maze (S2) environments (C0/C1 denote the context/state units, SEL the decision unit, for other abbreviations see Figure 3). Activation values are illustrated as vertical black lines whose height corresponds to the represented value (for each time step). All unit activation values lie between 0 (black dot) and 1 (full height black line). A filled circle for each time step indicates the agent's position. To simplify the matching of the agent's position/behavior and its network activation values the trajectories and the time lines have been segmented into five equally long time segments (labeled 0 to 4).

more detailed quantitative evaluation the reader is referred to Thieme (2002).

4.1 Simple T-mazes

In the first simple T-maze the network actually never activated the decision unit, and thus it remained purely reactive, always using the same sensorimotor mapping. Nevertheless, it solved the task by following the left wall to the goal when the light appeared on the left, and by turning right when facing the far wall of the junction in the other case. Similar, purely reactive solutions were reported by Bergfeldt and Linåker (2002).

The behavior of the agent in the second simple T-maze is illustrated in Figure 5. In the first case (a), where the light source appeared on the left, the agent moved all the way to the goal using the same sensorimotor mapping, with the appropriate right-turning behavior triggered by the far side of the junction. In the second case (b), while passing the light source, the activation of context unit C1 slowly went up over a few time steps until finally the decision unit (SEL) became active. The network then switched to a new sensorimotor mapping, providing the agent with left-wall-following behavior that took it all the way to the goal. It could be noted that in this case the decision unit, somewhat unnecessarily, remains active all the way to the goal, but we will see other examples later where this is not the case.

The connection weight matrices realizing the two sensorimotor mappings mentioned above are shown in Figure 6. In case (a), the agent moved fairly straight forward due to the combined effects of a strong positive bias weight for the right motor (MR) and the effect of the proximity sensors. These had a mostly inhibitory influence on the MR, but a mostly excitatory influence on the left motor (ML), which had a weaker positive bias. The turn to the right was triggered mostly by the proximity sensors on the left, DLL and DL, which became highly activated when the agent came very close to the wall at the far side of the junction (cf. Figure 5a). This made the agent turn right through a strong inhibition of the MR (see the large negative weights between DLL/DL and MR in Figure 6a) and further activation of the ML (positive weights between DLL/DL and ML in Figure 6a). The whole behavioral sequence in case (a) remained uninfluenced by the sensing of light on the left. The left

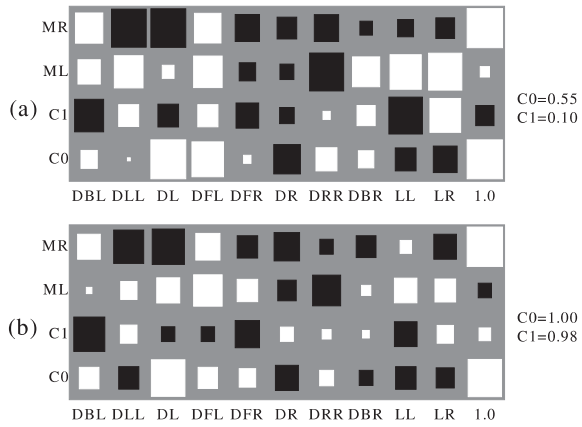

Figure 6 Hinton diagrams for the ESCN sensorimotor connection weights in the second simple T-maze (S2) environments. Connection weights are illustrated as black and white squares (black for negative/inhibitory weights, white for positive/excitatory weights), whose size reflects the weight magnitude. The columns labeled 1.0 contain the biases (or "bias weights"). C0 and C1 values to the right indicate the context unit activation values that led to the respective weight setting. For other abbreviations see Figure 3.

light sensor had a potential strong inhibitory influence on context unit C1 (large negative weight between LL and C1 in Figure 6a), but C1 had a negative bias anyway and thus remained more or less inactive throughout the whole sequence.

In case (b), on the other hand, the agent started off with the same initial sensorimotor mapping, but when encountering the light on the right side the right light sensor's (LR) activation also led to the activation of C1 (strong positive weight between LR and C1 in Figure 6a). After a short while context unit C1, as mentioned above, triggered the decision unit (SEL in Figure 5b) and this activated the context network, resulting in the new sensorimotor mapping illustrated in Figure 6b. The crucial difference was that contrary to the previous sensorimotor mapping (cf. Figure 6a), the ML now had a negative bias (cf. Figure 6b), and thus the agent had a greater tendency to turn left. The agent did not continue to move straight in the junction, but instead followed the left wall once the left proximity sensors DLL and DL's excitatory influence on the ML ceased. This is a relatively simple example of how slight modulation of the sensorimotor mapping, as embodied in the connection weights, can function as an STM mechanism.

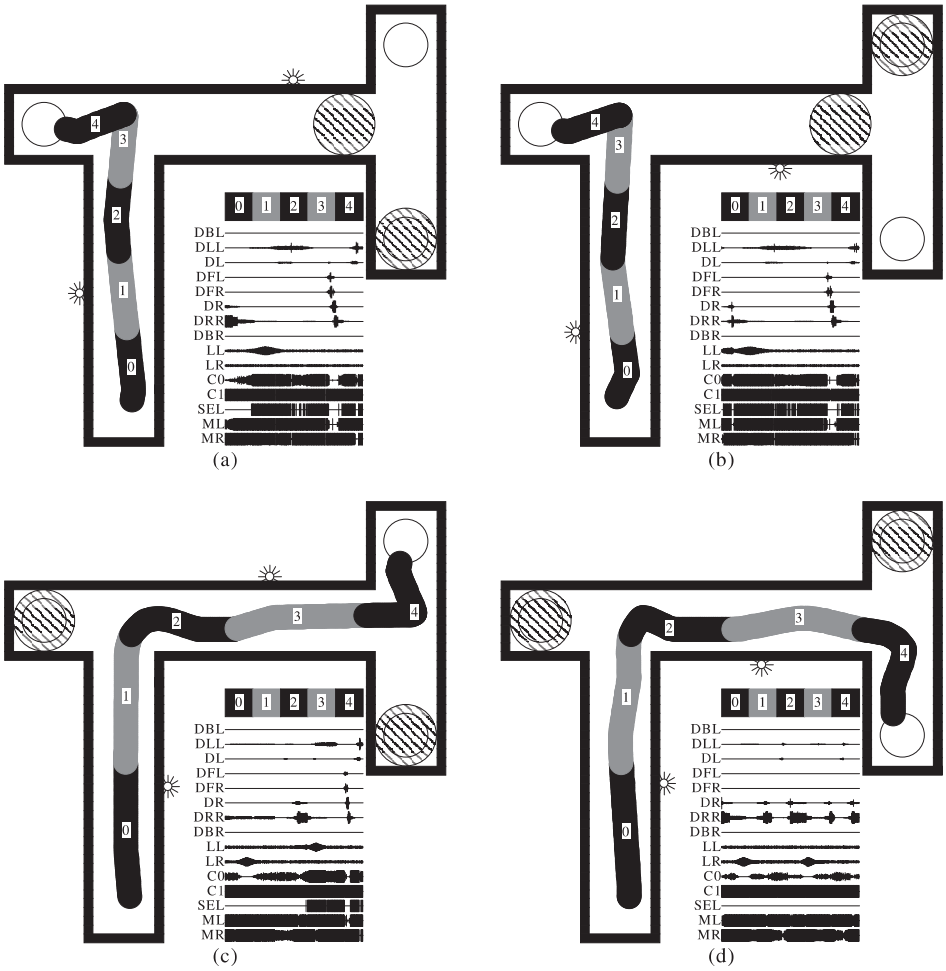

Figure 7 ESCN-controlled agents' behavior and activation values in the first repeated T-maze (R1) environments.

4.2 Repeated T-mazes

The behavior of the ESCN-controlled agents in the first repeated T-maze environments is illustrated in Figure 7. In case (d), where both light sources appeared on the agent's right side, the decision unit (SEL) never became active, that is, the agent solved the task in a purely reactive fashion. Hence, the network used the same sensorimotor mapping, following the right wall all the way to the goal. In the other situations, whenever the agent passed a light source on its left side the decision unit was activated, which led to a new sensorimotor mapping and different set of behaviors. The agent ceased to follow the wall and instead moved fairly straight forward until it faced the wall at the far side of the next junction where the appropriate left-turning behavior was engaged. Behaviorally this is very similar to the above solutions of the second simple T-maze environments. One type of light/turn situation is handled with a default wall-following behavior, whereas the other is handled by switching to a different set of behaviors, moving straight ahead and turning the opposite way in the junction.

The connection weight matrices underlying these behaviors are shown in Figure 8. Having analyzed a similar mechanism in detail above, the differences between the two weight settings, and their behavioral consequences, are relatively easy to explain here. The default sensorimotor mapping used initially in all four situations, and all the way to the goal in the fourth one (cf. Figure 7d), is realized by the weights illustrated in Figure 8a. A large positive bias for the ML and a very small negative bias for the MR give the agent a right-turn-, and thus right-wall-following tendency. When

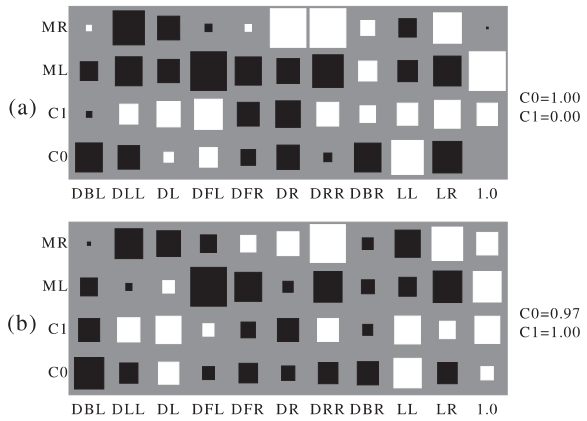

Figure 8 Hinton diagrams for the ESCN sensorimotor connection weights in the first repeated T-maze (R1) environments.

the agent encountered a light source on its left side this activated context unit C0 (C1 is active all the time), through the large positive weight between LL and C0 (cf. Figure 8a). When C0 was strongly activated, then the decision unit (SEL) became active as well (cf. Figure 7) and this led to the transition from the first sensorimotor mapping (cf. Figure 8a) to the second one, illustrated in Figure 8b. The crucial difference between the two seems to be the fact that in the second case both motors have roughly similarly large positive biases. This made the agent's default behavior straight forward motion rather than right turning (at the junction), and the turn to the left was facilitated by the right proximity sensors' inhibitory influence on the ML (negative weights between DFR/DR/DRR and ML in Figure 8b).

In the second repeated T-maze (not included due to lack of space) the network simply kept the initial sensorimotor mapping in all cases where the agent already reached the goal by turning left in the first junction. If the first light source instead appeared to the right, it switched to a second mapping that made the agent turn right at the first junction. Already when encountering the wall in that first junction it switched to a third mapping, and even to a fourth one if the second light was encountered on the left side.

4.3 Multiple T-mazes

In the first multiple T-maze[2] the initial sensorimotor mapping provided the agent with a right-wall-following tendency and thus took the agent all the way to the goal in a purely reactive fashion when lights were encountered only on the right side. When a light source was encountered on the left side, on the other hand, independent of whether this was the first or second light source, the agent switched to a sensorimotor mapping leading to a slight turn to the right, consequently moving through the corridor somewhat diagonally. As a result, the angle at which the agent would encounter the wall at the far side of the first junction depended on where the turn had been initiated. Hence, that first junction could be handled depending on which front proximity sensor was activated first/most. At the second junction, the agent could then simply always turn left since the above default strategy already covered situations requiring a second right turn.

The behavior of ESCN-controlled agents in the second multiple T-maze environments is illustrated in Figure 9. In this case, left turns were facilitated by left-hand wall-following, whereas right turns were achieved, as in several of the above cases, by first moving straight in the junction and then turning when facing the wall at the far side of the junction. It might be worth pointing out that the use of the decision unit (SEL), and consequently the use of the context network for modulation of sensorimotor weights in the function network, was highly selective and only occurred once in each scenario.

The ESCN's sensorimotor (or function network) connection weight settings used in the second multiple T-maze tasks are illustrated in Figure 10. The initial weight configuration (cf. Figure 10a) was such that in the absence of other stimuli, the approximately equally large positive motor biases resulted in straight forward motion. In the cases where the first light source was on the left side (cf. Figures 9a and 9b) the agent initiated a right-circling behavior, facilitated through the large negative weight between the LL and the MR, combined with a weaker inhibitory connection between LL and ML. This temporarily made the agent move away from the first junction. When facing the wall in the dead end this led to activation of the left frontal proximity sensor (DFL), which in turn briefly activated context unit C1 through the large positive weight between DFL and C1 (cf. Figure 10a), which in turn activated the decision unit.

As a result, the sensorimotor mapping was re-set to the one shown in Figure 10b, which realized a left-hand wall-following behavior. Similar to other cases discussed above, this was facilitated through a large

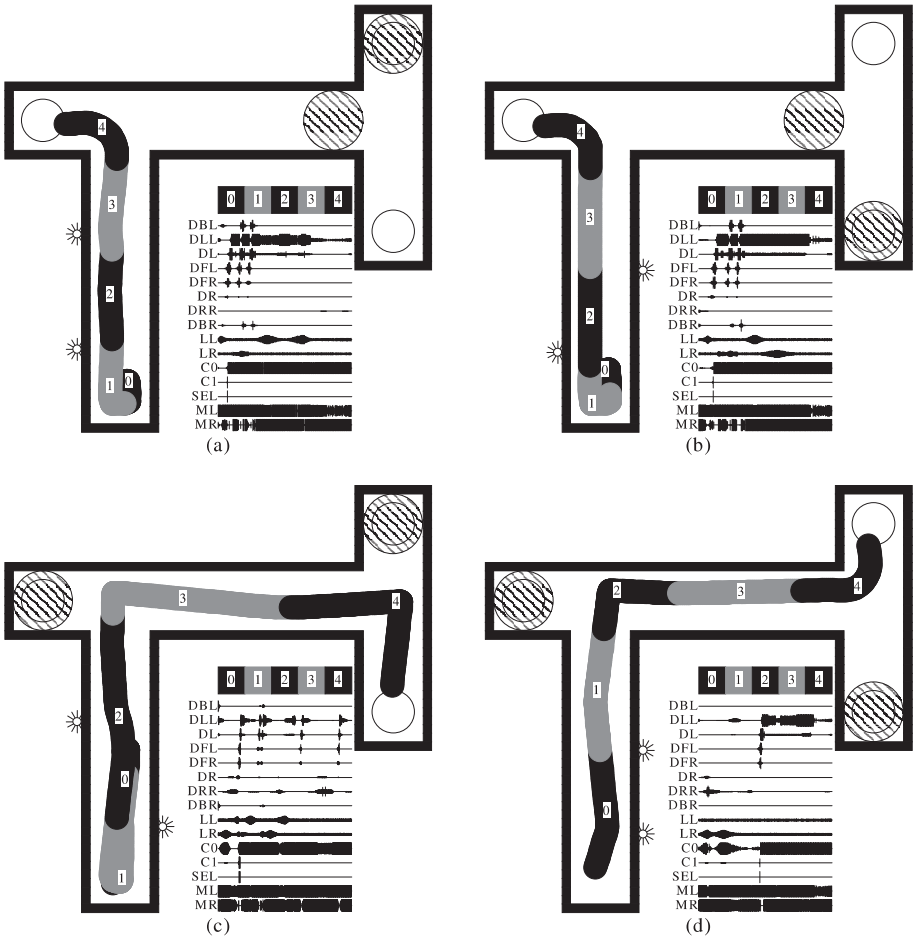

Figure 9 ESCN-controlled agents' behavior and activation values in the second multiple T-maze (M2) environments.

positive bias for the MR and a small negative bias for the ML, giving the agent a tendency to turn left in the absence of stimuli, combined with excitatory connections between the proximity sensors on the left (DLL, DL, DFL) and the ML. Hence, the agent would move forward as long as there was a wall to the left, and turn left as soon as that wall "disappeared" in the junction. In the cases illustrated in Figures 9a and 9b this took the agent all the way to the goal, in both cases more or less unaffected by the second light source on the left and right, respectively.

In the case illustrated in Figure 9c the agent encountered the first light source on its right-hand side, which also activated context unit C0 temporarily, through the large positive weight between left light sensor LL and C0 (cf. Figure 10a), but the behavior was not influenced at all. When sensing the second light source to the left, the agent, as in the above cases, turned around to move temporarily away from the first junction. During that turn the agent sensed the first light source again with the LR, which also activated context unit C0 again. Now, however, context unit C1 was activated at the same time, when facing the wall during the turn, through the large positive weight between left frontal proximity sensor DFL and C1 (cf. Figure 10a). The joint activation of both context units also triggered the decision unit (SEL), which resulted in the new sensorimotor weight setting illustrated in Figure 10c. This sensorimotor mapping made the agent move forward by default through two positive motor biases, and turn right in junctions. Both frontal proximity sensors, DFL and DFR, have excitatory connections to the ML and inhibitory connection to the MR (cf. Figure 10c). In this case these behaviors took the agent all the way to the goal, as illustrated in Figure 9c.

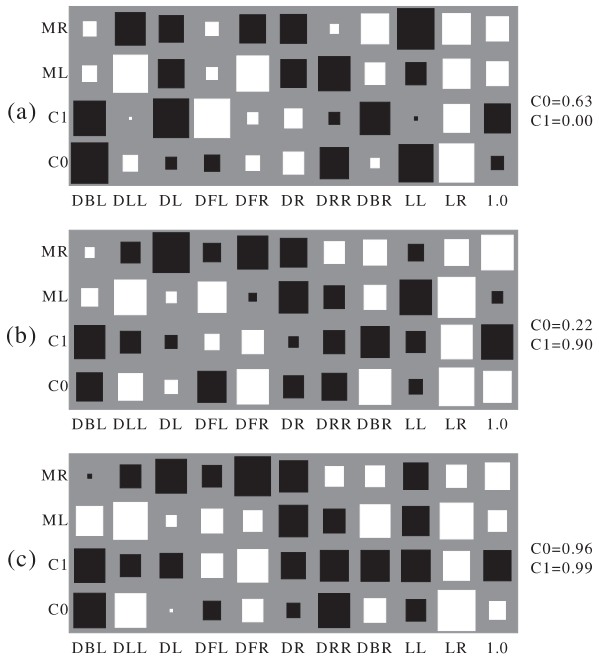

Figure 10 Hinton diagrams for the ESCN sensorimotor connection weights and resulting behaviors in the second multiple T-maze (M2) environments.

In the case where both light sources were placed on the right side, as illustrated in Figure 9d, the agent never turned away from the first junction. But as in the cases illustrated in Figure 9a and b, the first time the left front sensor DFL became activated, context unit C1 was activated temporarily as well. This in turn triggered the decision unit (SEL), which resulted in the sensorimotor mapping illustrated in Figure 10b, which even in this case took the agent to the goal through left-hand wall-following. Hence, the first two scenarios and the last one were all solved using the same combination of behaviors/sensorimotor mappings, but in the latter case the switch did not occur until after (or in) the first junction.

As mentioned briefly above, it should be noted that in all four cases the decision unit became active only once, that is, the sensorimotor mapping was adapted only once. Hence, the agent is basically controlled by purely reactive sensorimotor mappings all the time, except that at one point there is a switch from one reactive mechanism to another. It should be noted that at all other points in time the context unit activation values do not influence the sensorimotor mapping at all. This means STM is here not realized through the sustenance of neuronal activation patterns, but only through the dynamic modulation of reactive sensorimotor mappings, as embodied in the sensorimotor connection weights. Furthermore, it might be worth pointing out that in three out of four cases, the modulation was actually not triggered by the light stimuli. Hence, the new sensorimotor weights correspond to the behavior that will take the agent to the goal, but they could hardly be said to *represent* the light stimuli or the side on which they appeared, in the traditional sense of the term.

5 Summary

The main aim of this article has been to demonstrate and analyze the use of neuromodulation of sensorimotor mappings in higher-order recurrent neural robot controllers as an STM mechanism in delayed-response tasks. After some introductory discussion of the way feedback is used to provide STM in different types of recurrent neural nets, a number of experiments with extended sequential cascaded networks controlling simple robotic agents in six different delayed response tasks were presented. The analysis focused on the details of the internal workings of the ESCNs and their interaction with behavioral/environmental dynamics. The ESCN here serves to some degree as a representative of the class of higher-order recurrent neural networks, which has been used relatively little in autonomous agents and adaptive behavior research, and in other areas it has not been analyzed in the way it has been here.

The focus of our analysis has been on how STM is realized in ESCNs through synaptic plasticity and dynamic modulation of sensorimotor mappings (rather than through feedback of neuronal activation patterns, as in conventional recurrent nets), and how these internal dynamics interact with the external/behavioral dynamics, such as the use of reactive strategies. It has been demonstrated, in particular in the analysis of the last experimental scenario, how the ESCN, when at its best only makes very selective use of modulatory feedback, while as far as possible limiting itself to the use of reactive sensorimotor mechanisms and occasional switches between them.

It should be noted that the way STM is realized in these networks is very different from the traditional view of the neural mechanisms underlying STM in delayed response tasks. As discussed in the Introduction, the instruction stimulus has in many neuroscien-

tific experiments been shown to create or trigger neural activity that persists during the delay period and guides the delayed behavioral response afterwards. This neural correlate of STM is therefore commonly considered an internal *representation* of the original stimulus. The ESCN-controlled robots, on the other, as demonstrated in the previous section, clearly realize STM, but they do so in a way that is not representational in the traditional sense of a referential or correspondence notion of representation (cf., for example, Ziemke, 2001). This means that although the modulated sensorimotor mapping to some degree reflects the agent's history of interaction with the environment, in many cases there is no obvious correspondence between external stimuli and the internal parameters, which are anticipating future behavior rather than representing the past.

In summary, we believe that the alternative way of realizing STM in delayed response tasks that has been demonstrated in this article is relevant to cognitive- and neuroscientific theories and models of the relation between memory, representation and behavior. In future work, we intend to investigate further the highly selective use of feedback and the strong exploitation of environmental and behavioral dynamics, as exhibited in the last set of experiments, as well as address the question how a working balance between them is found in the processes of self-organization and agent–environment interaction.

Notes

1 Output and state units have biases (or bias weights); therefore 1 is appended to the input vector in these equations.

2 Not illustrated in detail, since the more interesting second multiple T-maze will be analyzed in detail below.

Acknowledgments

This work has been supported by a grant (1507/97) from the Knowledge Foundation, Stockholm, Sweden. The authors would like to thank Ezequiel Di Paolo and four anonymous reviewers for valuable comments on an earlier version of this article.

References

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*, 234–238.

Bergfeldt, N., & Linåker, F. (2002). Self-organized modulation of a neural robot controller. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2002)* (pp. 495–500). Piscataway, NJ: IEEE Press.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, *3*, 1184–1191.

Floreano, D., & Mondada, F. (1996). Evolution of plastic neurocontrollers for situated agents. In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, & S. Wilson (Eds.), *From animals to animats 4* (pp. 401–410). Cambridge, MA: MIT Press.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology*, *36*, 61–78.

Goldman-Rakic, P. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In F. Plum & V. Mountcastle (Eds.), *Handbook of Physiology* (pp. 373–417). Bethesda, MD: American Physiological Society.

Guigon, E., & Burnod, Y. (1995). Short-term memory. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 867–871). Cambridge, MA: MIT Press.

Husbands, P., Smith, T., Jakobi, N., & O'Shea, M. (1998). Better living through chemistry: Evolving GasNets for robot control. *Connection Science*, *10*(3–4), 185–210.

Ishiguro, A., Otsu, K., Fujii, A., Uchikawa, Y., Aoki, T., & Eggenberger, P. (2000). Evolving an adaptive controller for a legged-robot with dynamically-rearranging networks. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. Wilson (Eds.), *SAB2000 Proceedings Supplement* (pp. 235–244). Honolulu: The International Society for Adaptive Behavior.

Jakobi, N. (1997) Evolutionary robotics and the radical envelope of noise hypothesis. *Adaptive Behavior*, *6*(2), 325–368.

Linåker, F., & Jacobsson, H. (2001). Mobile robot learning of delayed response tasks through event extraction. In B. Nebel (Ed.), *Seventeenth International Joint Conference on AI (IJCAI)* (pp. 777–782). San Francisco, CA: Morgan Kaufmann.

Linåker, F., & Jacobsson, H. (2002). Learning delayed response tasks through unsupervised event extraction. *International Journal of Computational Intelligence and Applications*, *1*(4), 413–426.

Meeden, L. A. (1996). An incremental approach to developing intelligent neural network controllers for robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 26*(3), 474–485.

Meyrand, P., Simmers, J., & Moulins, M. (1991). Construction of a pattern-generating circuit with neurons of different networks. *Nature*, *351*, 60–63.

Michel, O. (1996) Khepera simulator package version 2.0. Freeware mobile robot simulator. Downloadable at http://diwww.epfl.ch/lami/team/michel/khep-sim/.

Mondada, F., Franzi, E., & Ienne, P. (1994). Mobile robot miniaturization: A tool for investigation in control algorithms. In T. Yoshikawa & F. Miyazaki (Eds.), *Proceedings of the 3rd International Symposium on Experimental Robotics* (pp. 501–513). Berlin: Springer.

Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics*. Cambridge, MA: MIT Press.

Pollack, J. B. (1987). Cascaded back-propagation on dynamic connectionist networks. In *Proceedings of the 9th Annual Conference of the Cognitive Science Society* (pp. 391–404). Cognitive Science Society.

Pollack, J. B. (1991). The induction of dynamical recognizers. *Machine Learning*, *7*, 227–252.

Rylatt, R. M., & Czarnecki, C. A. (1998). Beyond physical grounding and naive time: Investigations into short-term memory. In R. Pfeifer, B. Blumberg, J.-A. Meyer, & S. Wilson (Eds.), *From animals to animats 5* (pp. 22–31). Cambridge, MA: MIT Press.

Rylatt, R. M., & Czarnecki, C. A. (2000). Embedding connectionist autonomous agents in time: The 'road sign problem'. *Neural Processing Letters*, *12*, 145–158.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, *13*, 900–913.

Siegelmann, H. T. (1998). *Neural networks and analog computation: Beyond the Turing limit*. Boston, MA: Birkhäuser.

Siegelmann, H. T., & Sontag, E. D. (1995). On the computational power of neural nets, *Journal of Computer and System Sciences*, *50*(1), 132–150.

Thieme, M. (2002). *Intelligence without hesitation*. Master's dissertation HS-IDA-MD-02-001, Department of Computer Science, University of Skövde, Sweden.

Thieme, M., & Ziemke, T. (2002). The road sign problem revisited: Handling delayed response tasks with neural robot controllers. In B. Hallam, D. Floreano, J. Hallam, G. Hayes, & J.-A. Meyer (Eds.), *From animals to animats 7* (pp. 228–229). Cambridge, MA: MIT Press.

Ulbricht, C. (1996). Handling time-warped sequences with neural networks. In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, & S. Wilson (Eds.), *From Animals to Animats 4* (pp. 180–189). Cambridge, MA: MIT Press.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463.

Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature*, *382*, 629–632.

Ziemke, T. (1996). Towards adaptive behaviour system integration using connectionist infinite state automata. In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, & S. Wilson (Eds.), *From animals to animats 4* (pp. 145–154). Cambridge, MA: MIT Press.

Ziemke, T. (1999). Remembering how to behave: Recurrent neural networks for adaptive robot behavior. In L. R. Medsker & L. C. Jain (Eds.), *Recurrent neural networks: Design and applications* (pp. 355–389). New York: CRC Press.

Ziemke, T. (2000). On 'parts' and 'wholes' of adaptive behavior: Functional modularity and diachronic structure in recurrent neural robot controllers. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. Wilson, (eds.) *From animals to animats 6* (pp. 115–124). Cambridge, MA: MIT Press.

Ziemke, T. (2001). The construction of 'reality' in the robot. *Foundations of Science*, *6*(1), 163–233.

About the Authors

Tom Ziemke is Professor of Cognitive Science in the Department of Computer Science at the University of Skövde, Sweden. He received his doctorate from the University of Sheffield, UK, with a thesis on "Situated Neuro-Robotics and Interactive Cognition". His main research interests concern the mechanisms underlying enactive, embodied cognition. He is currently acting editor-in-chief of the *Connection Science* journal.

Mikael Thieme received a bachelor and a masters degree from the Department of Computer Science, University of Skövde, where until recently he also worked as lecturer.