Distributed Adaptive Control: Explorations in robotics and the biology of learning*

Paul F.M.J. Verschure

Institute for Neuroinformatics

ETH-UZ, Gloriastrasse 32, Zürich

CH 8006, Switzerland

www:http://www.ini.unizh.ch/~pfmjv

pfmjv@ini.phys.ethz.ch

Zürich, September 2, 1998

1 Introduction

Biological systems excel in their adaptive properties and the ability to develop appropriate behaviors to novel situations. In 1898 Thorndike reported one of the first systematic studies on animal learning [18]. A food deprived cat was placed in a so called puzzle box. In order to escape from the box and acquire food it had to manipulate particular aspects of this environment. Only through turning a button, pulling a string, depressing a lever, or pulling a wire loop would the escape latch be opened. Thorndike demonstrated that the time taken to escape from the box rapidly decreased over subsequent trails. In the field of robotics, despite the enormous advances in the technology used, no devices can be found which can match the behavior of the cats Thorndike studied 100 years ago. This can be interpreted as an invitation to reverse engineer nature. Such an approach has a long tradition. It is only recently, however, that important advances have been made in the study of these forms of behavior and their related brain mechanisms to make it a feasible option.

This paper will describe a synthetic approach towards the study of behavior, called Distributed Adaptive Control (DAC), which attempts to model these phenomena from a biological perspective. DAC has been developed using large scale computer simulations of neural models interfaced to real world devices. Before elaborating on DAC some definitions and clarifications on terminology will be provided.

2 Learning: definition and terminology

Although learning has been intensively studied over the last century no appropriate definition is available. In general, learning designates relatively long lasting changes in the

---

1

interaction between an organism and its environment. Hence, in the study of behavior operational definitions of learning, tied to specific kinds of experimental paradigms, are used. Before elaborating on the central components of the definition of learning a more fundamental question will be addressed; why do biological systems learn?

The present investigation assumes that learning expresses aspects of biological systems which allow them to deal with unpredictability [19]. Two types of unpredictability are distinguished: *somatic* and *environmental*. Somatic unpredictability results from the variability in the realization of the body plan. For instance, the detailed properties of the optics of an eye will vary over individuals. This property of biological systems constitutes a form of unpredictability which needs to be handled by a nervous system. For the present discussion we will make the assumption that the mechanisms which allow a nervous system to deal with somatic unpredictability are *developmental*. Environmental unpredictability is derived from the world in which biological systems have to function. This problem will be amplified by the extend to which the system under investigation is dependent on distal sensing, which is quite common for many organisms. The response of the organism to this source of unpredictability will be called *learning*.

Even though the distinction between developmental processes and those involved in learning is not very clear, for our present discussion we will only address the issue of learning. Learning will be operationally defined in terms of the standard paradigms used to investigate its properties; most notably classical and operant conditioning.

Classical, or Pavlovian, conditioning [15] refers to learning phenomena where initially neutral stimuli, or *conditioned stimuli* (CS), like lights and bells, are through their simultaneous presentation with motivational stimuli, *unconditioned stimuli* (US), like footshocks or food, able to trigger a *conditioned response* (CR). The success of this learning process is measured in terms of the probability of the occurrence of a CR after presentation of a CS. As to be expected the reality of animal behavior in the domain of classical conditioning is more complicated than was initially anticipated [10]. In order to place the discussed models in a proper context a number of additional properties of this type of learning need to be emphasized. Both at a behavioral and an anatomical level it is appropriate to distinguish consummatory, or specific, components from preparatory, or non-specific, ones [8, 9]. For instance, in the case of eyelid conditioning, where a tone (CS) is presented with an airpuff to the cornea (US), the animal will display a number of responses. Next to the closing of the eye lid, which can be seen as specific to the US, non-specific behavioral or autonomic responses can be observed; startle, freezing, withdrawal, changes in heartrate, breathing, or Galvanic skin response. The conditioned occurrence of these non-specific responses will follow a different temporal trajectory than the specific-responses. Non-specific responses show a fast acquisition (about 5 to 15 trials), while the development of the US specific CR takes a much larger number of trials. The CR will ultimately express the timing of the US. A more general interpretation of the behavior revealed in classical conditioning is that it allows behaving systems to learn about correlations between CS and US occurrences. To a certain extend one could speak of the substitution of the US by the CS through learning. This can be seen as a crude approximation of causal relationships in the world through correlative measures [22].

Operant, or instrumental, conditioning describes learning procedures in which the UR is contingent on a particular action displayed by the organism. The earlier mentioned puzzle box experiments of Thorndike can be taken as an example. The distinction between classical and operant conditioning is still debated in the field of animal learning. In the work presented here we make the assumption that both experimental paradigms address

complementary components of one *complete learning system.*

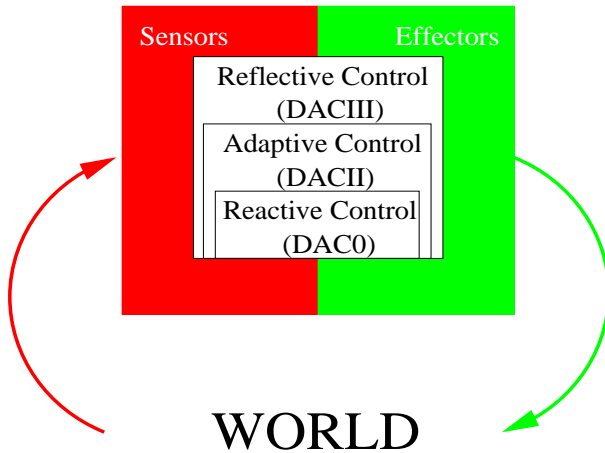# 3   Distributed Adaptive Control: The working hypotheses

Figure 1: The three levels of control distinguished in the design of a complete learning system.

The modeling series of Distributed Adaptive Control (DAC) [27] explores the question how biological systems acquire, retain, and express knowledge of their environment. In the evaluation of the different models the method of choice was the instantiation of simulated control structures in real-world devices (robots). This approach is seen as the instantiation of a research program of *synthetic epistemology.* In [21, 23, 24] the methodological and conceptual arguments for this choice are further elaborated.

The working hypotheses behind the DAC modeling series (illustrated in Figure 1) can be summarized as follows: First, the basic competence of a behaving system to effectively interact with its environment is derived from a *reactive control structure.* By solely relying on prewired relationships between US events and URs it will reflexively respond to immediate events. The triggering stimuli, USs, are derived from proximal sensors. URs can be given a more general interpretation as reflecting species specific behaviors. Second, as an *adaptive controller* the learning system will on one hand develop representations of states of the distal sensors which correlate with the events which activate the reactive controller. This element reflects aspects of the fast non-specific component of classical conditioning. On the other at this level of control reflexive responses can get further shaped, for instance in terms of their timing and duration. This aspect of adaptive control reflects the slow specific component of classical conditioning. Third, through forming more general representations of CS events and their relation to actions the learning system functions as a *reflective control structure.* At this level of control "plans" of actions can be formed through developing sequential representations. A system which comprises of these three components will be refered to as a *complete learning system.* It needs to be emphasized that the three levels of behavioral control are not defined as independent modules. As will be illustrated with the examples which will follow each level of control is strongly constrained by the preceding ones.

4 DACI and DACII: Evaluation of the hypothesis on the adaptive control structure

The DAC series started with a model proposed in [25] which was developed to solve some fundamental problems observed in models of so called *reinforcement learning* (i.e. [7, 17]), which are also widely used in the domain of machine learning (see [6] for a review). This class of models has proven to be very effective in dealing with a wide range of optimization problems and capture elements of classical conditioning. Their successful application to real-world problems, however, has still been rather limited. These models have two distinguishing features. In general, CSs are treated as predefined discrete sensory states. The actual learning process deals with the association of these predefined CS representations with particular responses, URs. The development of these associations are under the control of a global reinforcement signal derived from the presentation of a US. DAC, however, makes the assumption that an essential step in the learning process is the *acquisition* of CS representations based on the assumption of somatic and environmental unpredictability. In other words, the actual properties of the world defining a particular event have to be distinguished before any change to the behavior can be made. In addition, the global signals employed in order to convey an error measure is considered too strong and assumption. This does allow a formal treatment of these methods in terms of gradient descent methods. Despite their possible role as a heuristic in interpreting brain function [16], however, this assumption cannot be validated presently. DAC makes the more minimal assumption of the strict locality of the learning process.
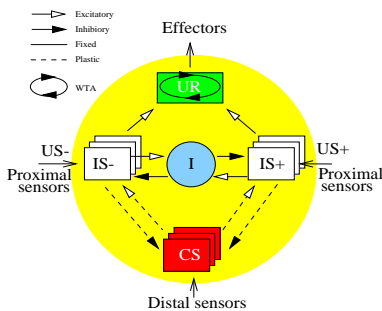

Figure 2: DACI and DACII: subsequent models of adaptive control. See text for explanation

In [25] it was shown that the assumption of the locality of learning could reliably reflect the acquisition of CS-US associations, using multiple CS modalities. In a next step, in order to find a behavioral validation of this proposal, a control structure was defined, called DACI, which was applied to a simulated robot in an obstacle avoidance and target approach task [27]. Figure 2 summarizes the basic components of this control structure. In the first experiments a single CS modality was considered derived from a range finder, while the USs were derived from collision (US$^-$) and an abstract target sensor (US$^+$). The reactive controller was defined by mapping the occurrence of an US onto a specific action of the robot, approach in case of an US$^+$ and avoidance in case of an US$^-$. An US event also induced activity in populations of units which reflect an *internal state* (IS); called *aversive* in case of an US$^-$ and *appetitive* in case of US$^+$. The projections of the proximal sensor to these IS populations conserved the topology present in the sensory sheet. Through the predefined interactions of the IS populations (indicated with I in Figure 2) preference relationships were established on order to resolve conflicts. Conflicts, for instance, occur in case the robot finds an obstacle in its path while following a gradient dispersed by a target. The units which ultimately drive the effectors receive inputs from the IS populations. Through a winner take all interaction one motor unit, and therefore action, is selected. The system would resort to its default behavior of moving forward, *exploration*, in case none of the IS populations were active. Learning proceeded by changing the connections between the IS

populations and the CS population. In this way associations between events transduced by the distal sensors and internal states could be acquired.

In these experiments it was demonstrated that DACI could successfully act in an envi-
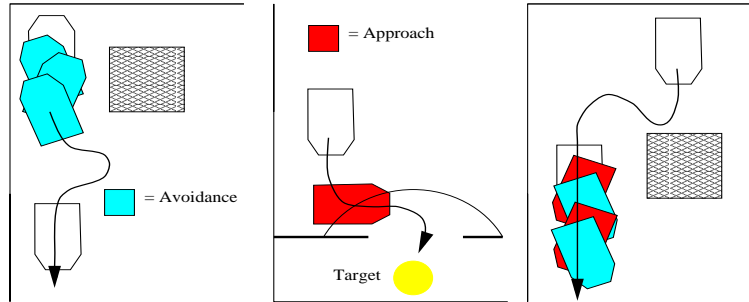

Figure 3: DACI: Emergent behaviors of the adaptive controller

ronment containing several obstacles and a single target. In addition, these experiments demonstrated the need for an activity dependent depression term in standard Hebbian learning rules. This rendered a learning rule similar to the well known Oja rule [13] which is able to extract the principal components from its input set. In this case this solution emerged out of the analysis of the performance of a behaving device. In addition these tests demonstrated that the relationship between observed behavioral regularities and the properties of the control structure can in certain cases only be understood in case the properties of the environment, the sensors and effectors of the behaving system, and the learning history are taken into account. Figure 3 shows some forms of structured behavior displayed by the adaptive controller, such as wall-following and impasse resolution, observed in these experiments. These behavioral patterns, however, are not explicitly represented by the control structure. The adaptive controller is only able to avoid or approach in response to immediate sensory events. It does not have the means to represent sequences of actions.

In [1] the optimization technique of genetic algorithms was used to demonstrate the stability of DACI over a wide range of parameters. A next series of tests involved the comparison and validation of the simulation results using real-robots [12]. In this case it was shown that the learning properties of DACI were independent of the actual distal sensors used. In [30] DACI was further generalized to a larger mobile platform applied to a visually guided block sorting task. In these experiments explicit performance comparisons were made between different proposals on rules governing synaptic plasticity. It was shown that so called "value based" learning rules [3] perform rather poorly in these types of tasks as compared to the learning method used in DAC.

In subsequent work [28] it was shown that the learning rule employed is prone to overgeneralize the representations it acquires. This is a fundamental problem for any local learning rule. Early CS-IS associations, expressed in the strength of the connections between the respective populations, in many cases would dominate later classifications. In the earlier described example of wall following these properties of the learning rule could induce an overgeneralization of this behavior; circling around obstacles. It was subsequently demonstrated that this problem could be solved while adhering to the assumption of the locality of learning process. Next to the feedforward path from the CS populations to the IS populations a recurrent inhibitory pathway between these populations was introduced. This extension was called DACII. As opposed to DACI, DACII would modify

the strength of the connections dependent on the deviation between actual CS events, transduced by the distal sensor, and CS events predicted through the IS activity. In [29] it is demonstrated that this extension to the original learning method, called *predictive Hebbian learning* provided the means to reliably capture CS representations and maintain these representations over extended periods of time in both simulated and real robots.

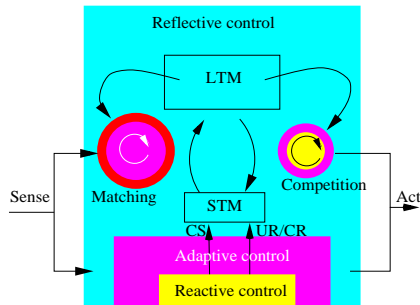# 5   DACIII: Evaluation of the hypothesis on the complete learning system

Figure 4: DACIII: an approximation of a complete learning system. See text for explanation

DACIII [20] was defined in order to explore the properties of reflective control. It constituted a first step towards the specification of a complete learning system. In [29] a more elaborate description of DACIII is provided, including evaluations using both simulated and real robots. The adaptive controller used in DACIII is similar to DAC II. The reflective controller modeled by DACIII consists of several components (see Figure 4). The first component is a transient memory buffer (Short Term Memory, STM) in which CS representations and their associated CRs are stored. Storage is conditional on the activity of an IS population. CS and CR representations are provided by the adaptive control structure. Each segment of STM contains one CS-CR pair. Conditional to the occurrence of a rewarding event, such as finding a target, the contents of STM are stored in a permanent Long Term Memory (LTM). CS representations, stored in LTM, are *matched* in parallel against ongoing sensory events. Matching LTM segments engage in a winner take all *competition*. The winning LTM segment triggers the next action, and reinserts itself in the STM buffer. In addition it will enhance the probability that the subsequent segment will match with future sensory events, in order to achieve chaining of subsequent LTM segments. This enhancement, however, is only transient.

The performance of DACIII was compared to that of DACII using a simulated robot similar to the one used in the first evaluation of DACI. In this comparison between DACII and DACIII the environment contained four targets each placed in a corner of a secluded space (Figure 5A). The gradient they dispersed only persisted over a short range. When a target is found by the robot it disappears from the environment. A new target will reappear in its place as soon as another target is found. The system could explore this environment for 10000 time step. The gradients dispersed by the targets were removed after 7000 time steps. After this time the system could find targets through either coincidence or the use of acquired behaviors. In this comparison DACII found 34 targets while accumulating 106 collisions. For DACIII this score was 53 and 73 respectively. The traces of the subsequent locations visited reflects the behavior of DACII and DACIII during the period that the target gradient was absent. The most salient aspect of the behavior of DACIII is its ability to acquire a stable behavioral pattern (Figure 5C). DACII, does not display this type of structuring of its behavior and visits all locations in the environment. These experiments

demonstrated that the complete learning system as implemented by DACIII is able to first, *acquire* appropriate representations of CS events, *retain* sequential representations of CS-CR couplings, and *express* these representations in order to successfully negotiate its environment.
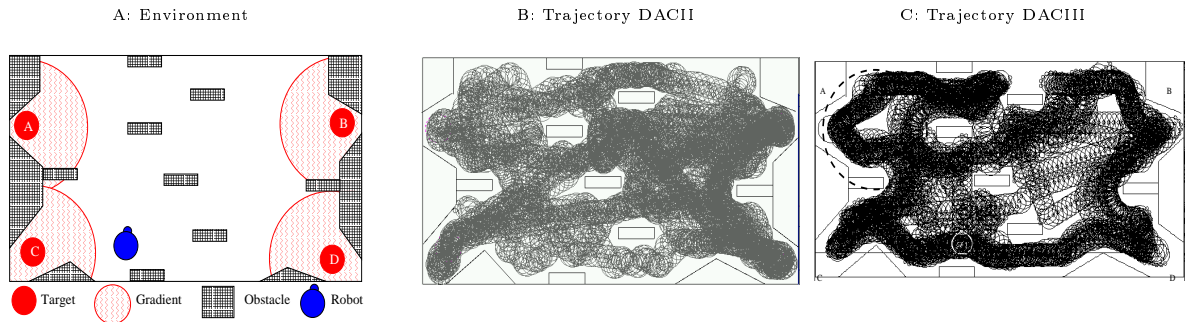
A: Environment   B: Trajectory DACII   C: Trajectory DACIII

Figure 5: DAC: A comparison of DACII and DACIII during the recall period.

6 Conclusion

The DAC series of models is by no means complete. For instance the properties of the specific learning system, and many elements of a reflective learning system have been neglected and are in the focus of our present activities. So far, however, DAC has shown that the basic assumptions behind this program suffice to define adaptive control structures which can capture aspects of advanced forms of behavior using strictly bottom-up principles. DAC has established a bridge between both the behavioral paradigms used in the study of learning and the neuroscientific explorations of these phenomena. In addition, it has made a number of suggestions which have shown to be of relevance to the fields of robotics [14], ethology [11], artificial intelligence [2, 4], and cognitive science [23]. In our current research of DAC we have focused on two additional themes. On one hand the development of more appropriate technology to construct behaving devices, especially through the use of more biologically realistic distal sensors (neuromorphic retinae) [5]. On the other more biologically realistic models of the brain mechanisms involved in the processing of distal sensors have been developed [26]. These efforts were undertaken in an attempt to facilitate a further validation of DAC at both the behavioral and the neuroscientific levels.

References

[1] N. Almassy and P. F. M. J. Verschure. Optimizing self-organizing control architectures with genetic algorithms: The interaction between natural selection and ontogenesis. In R. Manner and B. Manderick, editors, *Proceedings of the Second Conference on Parallel Problem Solving from Nature*, pages 451–460. Amsterdam: Elsevier, 1992.

[2] W.J. Clancey. *Situated Cognition: On human knowledge and computer representations*. Cambridge: Cambridge University Press, 1996.

[3] G.M. Edelman, G.N. Reeke, W.E. Gall, G. Tononi, D. Williams, and O. Sporns. Synthetic neural modeling applied to a real-world artifact. *Proceedings of the National Academy of Sciences of the USA*, 89:7267–7271, 1992.

[4] H. Hendriks-Jansen. *Catching Ourselves in the Act*. Cambridge Ma.: MIT press, 1996.

[5] G. Indeveri and P. F. M. J. Verschure. Autonomous vehicle guidance using analog VLSI neuromorphic sensors. In M. Hasler W. Gerstner, A. Germond and J-D. Nicoud, editors, *Proceedings Artificial Neural Networks-ICANN97: Lausanne, Switzerland*, pages 811–816. Lecture Notes in Computer Science. Berlin: Springer, 1997.

[6] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[7] A.H. Klopf. *The Hedonistic Neuron: A theory of memory, learning and intelligence*. Washington D.C.: Hemisphere, 1982.

[8] J. Konorski. *Integrative Activity of the Brain*. Chicago: University of Chicago Press, 1967.

[9] D. G. Lavond, Kim J. J., and Thompson R. F. Mammalian brain substrates of aversive classical conditioning. *Annual Review of Psychology*, 44:317–342, 1993.

[10] N.J. Mackintosh. *The Psychology of Animal Learning*. New York: Academic Press, 1972.

[11] D. McFarland and T. Bosser. *Intelligent Behavior in Animals and Robots*. Cambridge, Ma.: MIT press, 1993.

[12] F. Mondada and P. F. M. J. Verschure. Modeling system-environment interaction: The complementary roles of simulations and real world artifacts. In , editor, *Proceedings of the Second European Conference on Artificial Life*, pages 808–817. Cambridge, Ma.: MIT press, 1993.

[13] E. Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.

[14] O. Omidvar and P. Van der Smagt. *Neural Systems for Robotics*. New York: Academic Press, 1997.

[15] I. P. Pavlov. *Conditioned Reflexes*. Oxford: Oxford University Press, 1927.

[16] W. Schultz, P. Dayan, and P.R. Montague. A neural substrate of prediction and reward. *Science*, 275:1593–1599, 1997.

[17] R. S. Sutton and A. G. Barto. Toward a modern theory of adaptive networks: expectations and prediction. *Psychological Review*, 88:135–170, 1981.

[18] E.L. Thorndike. *Animal Intelligence: an experimental study of the associative processes in animals*. Number 8 in The Psychological Review Series of Monograph Supplements. New York: Macmillan, 1898.

[19] P. F. M. J. Verschure. 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*, pages 653–658. Hillsdale, N.J.: Erlbaum, 1992.

[20] P. F. M. J. Verschure. The cognitive development of an autonomous behaving artifact: The self-organization of categorization, sequencing, and chunking. In H. Cruze, H. Ritter, and J. Dean, editors, *Proceedings of Prerational Intelligence*, pages 95–117. Bielefeld: ZiF, 1993.

[21] P. F. M. J. Verschure. Formal minds and biological brains. *IEEE expert*, 8(5):66–75, 1993.

[22] P. F. M. J. Verschure. Minds, brains, and robots: Explorations in distributed adaptive control. In A.B. Soares, editor, *Proceedings of the Second Brazilian-International Conference on Cognitive Science*, 1996.

[23] P. F. M. J. Verschure. Connectionist explanation: Taking positions in the mind-brain dilemma. In G. Dorffner, editor, *Neural Networks and a New Artificial Intelligence*, pages 133–188. London: Thompson, 1997. First presented at ZiF workshop Mind and Brain, 1990.

[24] P. F. M. J. Verschure. Synthetic epistemology: The acquisition, retention, and expression of knowledge in natural and synthetic systems. In *Proceedings World Conference on Computational Intelligence 1998, Anchorage*, pages 147–153. IEEE, 1998.

[25] P. F. M. J. Verschure and A. C. C. Coolen. Adaptive fields: Distributed representations of classically conditioned associations. *Network*, 2:189–206, 1991.

[26] P. F. M. J. Verschure and P. König. Modulation of temporal interactions in cortical circuits. In H-M Gross, editor, *Proceedings of SOAVE97*, pages 77–88. Dusseldorf: DVI, 1997.

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

[28] P. F. M. J. Verschure and R. Pfeifer. Categorization, representations, and the dynamics of system-environment interaction: a case study in autonomous systems. In J. A. Meyer, H. Roitblat, and S. Wilson, editors, *From Animals to Animats: Proceedings of the Second International Conference on Simulation of Adaptive behavior. Honolulu: Hawaii*, pages 210–217. Cambridge, Ma.: MIT press, 1992.

[29] P. F. M. J. Verschure and T. Voegtlin. A bottom-up approach towards the acquisition, retention, and expression of sequential representations: Distributed adaptive control III. *Neural Networks*, page In Press, 1998.

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