Synthetic Epistemology: The acquisition, retention, and expression of knowledge in natural and synthetic systems

Paul F.M.J. Verschure

Abstract— **The multi disciplinary study of mind, brain, and behavior has reached a point where we seem to be confronted with the dilemma of studying either the computational mind or the dynamical brain. The cognitivist study of the computational mind has shown to be very effective in describing many elements of "high level" cognition. In the mean time, however, its most explicit expression, classical AI, is facing a number of fundamental problems. This has created a situation where arguments are raised in favor of abandoning the computational view. In this paper this dilemma is described. It is demonstrated that only one problem is hidden behind the multiple challenges facing the traditional computational program; the problem of a prioris. It is shown that the central component in a solution of this problem focuses on the nature of the knowledge ascribed or implemented by cognitive systems. A research program is defined, synthetic epistemology, which aims at addressing this question. As an example of this approach the modeling series of Distributed Adaptive Control is described. It is shown that this modeling series, which reflects elements of the behavioral paradigms of classical and operant conditioning and has been shown to be consistent with aspects of the correlated neural substrate, provides an alternative unifying perspective to the dilemma facing the study of mind, brain, and behavior.**

Keywords— **Mind, brain, behavior, Distributed Adaptive Control, learning, neural model, epistemology, robot, conditioning, artificial intelligence.**

I. Introduction

The study of mind, brain, and behavior, has united itself in the last decades in a so called cognitive science. This conglomerate of disciplines, spanning from neuroscience to anthropology, expresses that the study of the mind and brain needs to proceed over a wide multi-disciplinary front. The relatively recent field of artificial intelligence played a novel and central role in this endeavor. Through the construction of working computer programs it would provide a synthetic evaluation of the proposed theories of cognition. An important element in the research strategy of cognitive science was the emphasis on variables internal to the behaving system, i.e. knowledge, and the reliance on a computer metaphor. The main successes of this approach have been in the domain of "higher cognition", for instance problem solving and language. Even though it has always known criticism only more recently a number of fundamental problems have been identified. This development has spawned alternative proposals towards the study of mind, brain, and behavior. A central feature of these is that they dismiss many of the central assumptions of the preceding one, which is quite common in case a novel paradigm needs to define its own identity.

In this paper this development will be further analyzed. The traditional paradigm of cognitive science, with its emphasis of the computational mind, will be shortly defined and its main problems summarized. It will be argued that these identified problems are the expression of a single underlying theme; *the problem of a prioris.* This analysis provides the background for the description of an alternative experimental approach towards the study of mind, brain, and behavior, called synthetic epistemology. After a description of its central conceptual and methodological features a modeling series, Distributed Adaptive Control, will be shortly described demonstrating that a consistent approach towards realizing the wish of a cognitive science, which incorporates elements of the dynamical brain and the computational mind, is possible.

II. The computational mind

The dominant paradigm in the study of mind, brain, and behavior can be called symbolic cognitive psychology [1]. This approach bases its explanations of cognition on a so called *knowledge level.* A central principle in a knowledge level explanation is the law of rationality: a rational system will use its knowledge in order to reach its goals. A paradigmatic example of this approach, which constituted the core of the artificial intelligence program, is the hypothesis of *Physical Symbol Systems* (PSS) put forward by Newell and Simon [2]. The states of the environment, in which a PSS is embedded, are transduced to internal symbolic representations. A PSS is able to perform operations on these representations, the result of which are again symbolic expressions. The output of the system is defined through the interpretation of symbolic expressions. The symbols and operators, which define a PSS, form a finite set of logical axioms which can be seen as its *world model.* This approach is strongly inspired by the digital computer. It is emphasized, however, that it should not be seen as the application of a computer metaphor, but as a scientific theory on the structure of the mind. A classical example of this proposal is the General Problem Solver (GPS) [3]. GPS was mainly applied to the solution of logical puzzles. It created quite some excitement, however, when it was shown that it displayed problem solving behavior which seemed quite similar to that of humans exposed to the same task. Recently a further step in the same tradition, SOAR, was proposed as a unifying theory of cognition (see [4] for a

Institute of Neuroinformatics, ETH-UZ, Gloriastrasse 32, CH-8006, Zürich, Switzerland. pfmjv@ini.phys.ethz.ch

review). An interesting feature of SOAR is that it provides a more dynamic approach towards problem solving. Even though SOAR still works with a predefined world model it can expand its production system by means of *chuncking*. In case a provided problem leads to an impasse, i.e. a production rule to solve it is not directly available, a so called subgoal is defined and the impasse becomes a subproblem space. SOAR will cycle through its normal decision cycle, applying operator after operator, until the subgoal state is reached. After the resolution of this impasse the initial state of the subproblem space and the successfully applied operator are chunked into a new operator. Chuncking prevents SOAR to solve the same problem twice and will optimize its performance.

The rationalistic philosophy behind the computationalist program has a number of interesting implications which are relevant to the present discussion. By emphasizing the rules and representations implementing the rational mind, the physical properties of the substrate of implementation is of no relevance. Although this problem of *multi instantiation* seems unpleasant from the perspective of a unified view relating the mind to the brain, the commitment to a computational program defines a special science of the disembodied mind which proceeds in isolation from the natural sciences [5].

Next to the implication of multi-instantiation the cognitivist program leads to a strict nativism. For instance, the explanation of cognition provided by a PSS is crucially dependent on the a priori definition of its world model. One has to assume that the world reveals itself to a behaving system in terms of discrete events that are all a priorily represented in terms of discrete symbols, transduction rules and appropriate operators. In the discussion on the nature of linguistic processes, which has strongly developed along cognitivist terms, this has induced the believe of a “language organ” which is fully genetically predefined [6]. The more fundamental argument behind this nativist position is that learning is simply not possible since learning through induction still requires a priori hypothesis (see [7] for the most explicit rendering of this argument). The earlier described mechanism of chuncking of SOAR provides a clear illustration. Even though SOAR can “learn” new operators through chuncking, the building blocks of a chunk are necessarily derived from the symbols and operators present in the predefined world model. The system can in principle never escape the logical closure of its world model. Remaining close to its computer metaphor this is, however, considered not a bug but a feature.

III. The problem of a prioris

Over the last years a number of fundamental problems of the computational program have been identified; the frame problem [8], the symbol grounding problem [9], [10], the frame of reference problem [11], and the problem of situatedness [12]. These problems have been extensively discussed in previous work [13], [14] they, however, can be seen as the result of one underlying issue, called *the problem of a prioris* [15]. In short this problem is created by the critical dependence of a model of a cognitive process on the a priori specification of a world model. As a result such an approach runs the risk of specifying a system which is not grounded in the world in which it finds itself, is prone to mix up the domain ontologies involved, relies on a representational granularity which induces a search problem, and ignores some pertinent elements of system-environment interaction. In order to illustrate how each of the challenges facing the computational program can be seen as a symptom of the problem of a prioris they will be shortly described.

The frame problem relates to the issue of search, central to a PSS. A somewhat interesting cognitive system will have to deal with an extensive world model. Every time something changes in the environment these internal representations need to be reevaluated. The problem is, however, that this update has to be carried out on the complete world model. This will lead to a strong deterioration of the capability of the system to act, and in most cases to its ungraceful end. The problem of a prioris shows up in the frame problem since the assumption was made that representations are discrete; i.e. one symbol stands for one particular event in the world. This choice of the granularity and mutual independence of internal representations automatically places the full burden of ”computation” on search. Without any constraints on this search process the frame problem has to appear. The symbol grounding problem addresses the issue of how meaning is assigned to symbols. A knowledge level explanation does not provide any insight into this issue, a PSS is a pure syntactic system. It assumes the existence of rules and representations. This relates to the frame of reference problem, which reemphasizes the importance of being clear about the domain ontologies involved in describing or defining a cognitive system; i.e. designer, observer, or behaving system. For instance, in case of constructed systems the domain ontology will be provided by its designers and be grounded in their domain ontology. In psychology it will be the experimenter who will attribute knowledge and goals to the subject. The requirement of a full a priori specification will imply that in most cases there will be a mismatch between this attributed world model and the actual world in which the behaving system finds itself. The issue of situatedness further refines the possible sources of this mismatch. Behaving, and also cognitive, systems are physically instantiated and embedded in a dynamic and unpredictable world. Moreover, they have to perform under time pressure. The issue of situatedness also points out that a world model need not be constructed a priorily. The world can be sensed and as such function as its own model. A PSS, however, is not situated in the real world, it exists in the domain ontology of its designer.

IV. Connectionism and New AI

In the last decade cognitive science has seen a strong surge towards alternative models of cognition. These activities can be partly explained by the perceived problems of the dominating paradigm of symbolic cognitive psychology, which were shortly pointed out above. Approaches like connectionism [16] and "new AI" [17], [18] propose a biologically inspired program focusing on a more real-world oriented perspective. Connectionism emphasized the role of learning and distributed representations and seemed to provide a promising alternative paradigm, which attracted a lot of attention. The viability of these proposals can be measured in terms of how well they deal with the problem of a prioris. It has been shown previously, however, that paradigmatic examples of these supposed alternative approaches and their derivatives do not necessarily satisfy the criteria outlined above [19], [14], [20]. Also these proposals were shown to be critically dependent on their designers domain ontology. The surprising conclusion is that these alternative approaches have provided additional arguments in favor of extreme nativism. They were not able to relax to necessity of strict a priori specification.

An alternative perspective provided by so called "New AI" is that explanations of cognition should proceed without relying on the notion of representations [17] or goals [18]. In practical terms this means that one negates the principle of rationality. This assertion in itself, however, does not automatically solve the problem of a prioris. Moreover, despite its fundamental limitations the computational program has provided a conceptual framework for describing aspects of cognition which cannot be denied; i.e. problem solving and language. This seems to imply that one either denies the physical instantiation of cognitive systems and retains the disembodied mind or negates the principle of rationality and is left with a mindless body. Both scenarios seem unwanted.

The original question behind the computational program was how a behaving system *uses* its knowledge to achieve its goals. The previous argument has shown that the proposed answer to this question, in terms of PSS, got bogged down in a number of fundamental problems, captured in the problem of a prioris. This result leaves two options. One can insist on strict nativism, and its dissatisfactory implication that central components of psychological explanations are beyond scientific scrutiny. This option is a basic tenet of the hypothesis of the modularity of mind (Fodor, 1983). Alternatively one could argue that before we can address the issue how a system *uses* its knowledge the question of how this knowledge is *acquired* and *retained*, needs to be explored. It is only through answering this question of the acquisition and retention of knowledge that a more realistic view on the "world models" entertained by the only forms of cognition we know, biological systems, can be developed. Hence the program of pursuing an "artificial intelligence" which got bogged down in the problem of a prioris, needs to be replaced with a program of *synthetic epistemology.*

V. Synthetic epistemology

The assumption that the explanation of behavior requires an understanding of variables mediating between sensing and acting, which can be designated with the constructs representations and goals, has a long tradition not only in the research program of traditional AI, but also in psychology and neuroscience. The question, however, is how these variables can be defined in terms which allow the study of both the mind and the brain. For instance, in case we want to interpret the behavior of a system following the principle of rationality what are the possible structural correlates of knowledge and goals? The approach presented here makes the assumption that in order to pursue this question we need to rely on a synthetic approach using real-world devices. This assumption is based on two considerations. First, the problem of a prioris showed that by making too strong assumptions on the a priori properties of a world model a number of fundamental problem arise. This raises the question how a "world model" can be defined without falling victim to the problem of a prioris. In the present approach the assumption is made that given that the world is an unpredictable place, the theme of situatedness, world models cannot be fully pre specified but need to be acquired. Only in this way can the body of knowledge of a behaving system be grounded in its "experience"; its interaction with the environment. Hence, the use of knowledge can only be studied by means of systems that actually interact with the world through sensors and effectors. Second, in developing multi leveled scenarios on cognitive systems, which include both functional and structural components, the actual research methodology needs to be specified. The present proposal adheres to a methodology of *convergent validation* [14]. This method prescribes that the study of particular functional properties of a system needs to be simultaneously constrained at multiple levels of description. In the present case these levels of description include both the neural and the behavioral domains. The neural level can be approximated through simulation studies using a standard or a custom computational infrastructure. The behavioral level, however, requires that these models are embedded in real-world devices, since behavior is a real-world real-time phenomenon. The above arguments demonstrate that in case we want to pursue an integrated study of mind, brain, and behavior, we necessarily have to rely on synthetic methods, both for methodological and conceptual reasons.

VI. Distributed Adaptive Control

An example of the study of synthetic epistemology following the above outlined method of convergent validation is the modeling series of Distributed Adaptive Control (DAC) [21]. DAC focuses on the study of the behavioral paradigms of classical and operant conditioning (i.e [22]) from a neural perspective. DAC is based on a number of assumptions regarding the structuring of a nervous system. At least three levels of control need to be distinguished. First, by solely relying on prewired reflexive relationships between sensory events and actions the system functions as a *reactive controller* (DAC0). It will reflexively respond to immediate events. These reflexes are based on predefined relationships between events on proximal sensors and behavioral stereotypes. Second, as an *adaptive controller* (DACI-DACII) the system will develop representations of events that correlate with the activation of the reactive control structure. In addition the reflexive actions can be reshaped in order to better reflect the properties of an environmental perturbation. At the level of *reflective control* (DACIII) more extended representations of sensory events and motor actions will be formed, for instance expressing their relationship over time (see [15], [23] for a more extensive description).

The three levels of control distinguished do not function as independent modules but are closely coupled. For instance, a reactive control structure provides a behaving system with a basic competence such as reducing damage to the body. It, however, only responds to events occurring on the proximal sensors. The adaptive controller provides an interface between the reactive system and the distal sensors. At this stage representations of events at the distal sensors are acquired. What constitutes an event, however, depends on the properties of the reactive controller. A typical example is the occurrence of a collision, sensed through a proximity sensor, which induces an avoidance response. The adaptive controller will try to develop a representation of the distal sensor states which correlate with such an event: a sensory prototype. In our subsequent models of the adaptive controller, DACI and DACII, an important constraint was the insistence on local learning rules in acquiring these sensory prototypes [20]. This is in sharp contrast to the main stream of methods explored in the domain of machine learning (see [24] for a review). In [25] it was demonstrated that a local learning method can develop stable representations of distal sensor events provided it is embedded in an appropriate recurrent circuit (Figure 1A). This method was called *Predictive Hebbian Learning* [23].
In our model of the adaptive controller the sensory prototypes expressing distal sensor events are defined by the synaptic strength of the connections between populations of units which respond to the distal sensor and those responding to events of the proximal sensors. Proximal sensors activate units representing internal states, for instance "aversive" in case of collisions. Through the internal state units reflexive actions are triggered, "avoid". In [25] it is demonstrated that this circuit will develop very stable representations of distal sensor events conditional on the internal states of the system. A sensory prototype will therefore automatically relate a sensory event with particular behavioral stereotypes. In [20], [26] it is demonstrated how this control structure can deal with a large number of distal sensor domains. Predictive Hebbian Learning is further described in [23], [27].

An interesting feature of our model of adaptive control, which has been explored using both simulated and real robots using a wide range of distal and proximal sensors [28], [26], is that it gave rise to behavioral regularities which emerged out of the continuous interaction of the robot with the environment, such as wall following [21] (Figure 1B). During learning the system found itself parallel to walls while encountering a gradient dispersed by a target. This lead to an association between the distal sensor event, induced by the wall, and the internal state of

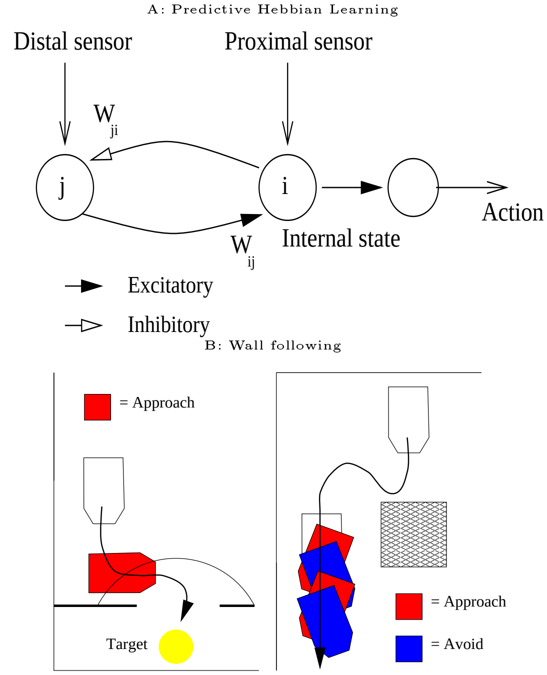

Fig. 1. Adaptive control-DACII. A: Predictive Hebbian learning. Unit i responds to the distal sensor, while unit j expresses an internal state triggered by a proximal sensor. Unit i excites unit j with synaptic strength w_{ij} while unit j inhibits units i with connection strength w_{ji}. Both the feedforward and the recurrent connections are updated following an Hebbian rule. B: Wall following an example of "emergent" behavior.

x

appetitive. This internal state was activated through the proximal sensors detecting the gradient and induced an approach action: turn in the direction of the gradient. Later in the trial this approach behavior was generalized to other situations where the system found itself parallel to walls. Approaching an object too closely, however, would induce an aversive state due to the previous experiences with collisions. Effectively wall following was constructed out of the alternation between approach actions and avoidance behavior.

VII. Reflective control

In a subsequent model, [29], [23], we addressed the issue of sequence learning as an approximation of reflective control (Figure 2A). This control structure, DACIII, builds up a representation of sensory prototypes and their related actions, as acquired and expressed by the adaptive control structure, in a transient short term memory (STM) buffer. Storage in STM is conditional on the activity of an internal state, aversive or appetitive. In case the system finds a target, "reward", the STM sequence of sense-act representations, segments, stored in STM are retained in a permanent long term memory (LTM). The sensory prototypes in each LTM segment will match with ongoing events on the distal sensor followed by a global winner take all competition between matching segments. The winning, or dominant, segment will induce its action and overrule the adaptive control structure. In addition the winning segment will reinsert itself in the STM buffer. In this way recombinations of different LTM segments or ongoing events at the adaptive control level can be formed. Chaining through the segments of a sequence is achieved through increasing the likelihood that the segment following a dominant segment will match future distal sensor events.

(Figure 2B) provides an example of the structuring of be-

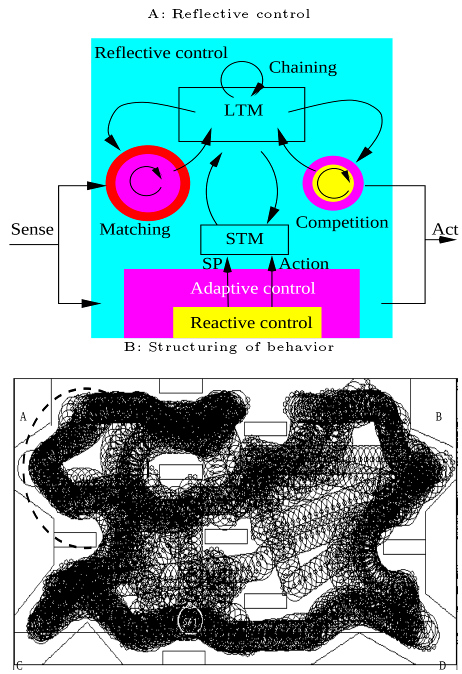

Fig. 2. Reflective control. A: The main components of reflective control. SP = sensory prototype. B: An example of the structuring of behavior through reflective control. In this simulation experiment the distal sensor was defined by a range finder while the proximity sensors were a collision sensor (aversive-avoid) and a target sensor (appetitive-approach). The environment consisted of four targets (A, B, C, D), which dispersed a gradient indicated by the dashed circle. In this trial, which lasted 3000 time steps, the target gradients were removed after 2000 time steps. In this way a *recall period* was defined. The trajectory displayed represents all positions visited by the simulated robot during the trial.

x

havior through reflective control. The densely labeled positions in the environment are visited at regular intervals during the trial. The system settled in a stable trajectory in a clockwise direction. It found a total of 25 targets and suffered 18 collisions. A reflective control structure which did not form recombinations had a similar number of collisions but only found 19 targets. The adaptive controller accumulated 32 collisions and found 9 targets. In [23] it is shown that the reflective control structure generalizes well to a microrobot, Khepera (K-team, Lausanne, Switzerland) using a color CCD camera, while [30] provides an extensive evaluation of the detailed performance differences.

VIII. A knowledge level interpretation

In [26], [15], [23] the relationship between the DAC series of models and the biology and psychology of learning and memory is described. In our present discussion, which focuses on the question whether the view of the computational mind can be reconciled with that of the dynamical brain, the emphasis is on the reinterpretation of a knowledge level description of behavior and the perspective of synthetic epistemology of which DAC is an expression.

A first question is the notion of representation. A PSS relies on the full a priori specification of symbolic representations and transduction rules of events in its environment. The model of the adaptive controller shows that this constraint can be relaxed and that prespecification of a reduced set of sense-act relationships suffices, only based on the relationships between proximal sensor events and behavioral stereotypes. Distal sensor representations can to a large extent be acquired. This implies that they are automatically grounded in the "experience" of the behaving system. The problem of search is partly addressed through the ability of the sensory prototypes to generalize to a large class of discrete environmental events [25].

The wall following behavior, Figure 1B, provides a vivid illustration of the frame of reference problem. Even though the adaptive control structure is not able to represent sequences it does display organized patterns of behavior. From the perspective of the outside observer this could be described in terms of the goal of the system to follow a wall, and assumptions about the rules it would use to achieve this goal. Analyzing the adaptive controller, however, shows that in the perspective of the control structure something rather different happens: turning towards surfaces which are parallel to its own orientation and subsequently turning away from surfaces in front of it. This demonstrates that not in all cases a functional dissection of behavior, using the knowledge level, is matched by the structural properties of the system which creates the observed behavior. There is no claim here that this constitutes an explanation of the way in which this behavior occurs in biological systems. This example does illustrate, however, that the importance of a synthetic approach towards the study of knowledge level questions. Only through the full experimental control of the real-world device (robot), the control structure, and the environment could the genesis of these behavioral patterns be explained in terms of the learning history, and the properties of the robot and the environment. This level of experimental control cannot be achieved using standard experimental methods.

Our model of the reflective controller, DACIII, is the closest approximation of a synthetic rational system, which uses its knowledge to reach its goals. The goals are defined in terms of its *internal states*, i.e. avoid or approach. The knowledge it brings to bear on reaching these goals are the acquired LTM segments, which can be interpreted as the world model of the system. This world model, however, is at no point in time fixed. The content of LTM can change at any time due to new experiences.

IX. Discussion

The traditional approach towards the study of the mind has proceeded in isolation from the study of the brain and placed itself in an extreme nativist position. This raises the question whether, as some propose, the computationalist program should be fully dismantled and replaced by an alternative such as "New AI". It was argued that since the study of the mind and brain is necessarily a multidisciplinary adventure an advance can only be expected in case a neutral specification of cognitive systems can be formulated which allows the construction of principles which can mediate between the different levels of description. As opposed to rejecting the traditional computational view, it should be taken as one important perspective which needs to be integrated with others such as a biological one. The DAC modeling series, which captures elements of the behavioral paradigms of classical and operant conditioning and has been shown to not violate basic elements of the neural substrate, illustrated that such a compromise is possible. DAC is the expression of an approach of synthetic epistemology which is defined as the study of the acquisition, retention, and expression of knowledge by biological systems. Given the nature of the questions posed, and the research methodology of convergent validation this approach is automatically of a synthetic nature. In order to further develop the tools which would facilitate such an exploration we are presently incoporating aVLSI, neuromorphic, sensors into our modile platforms [31].

Although many problems remain to be solved, and the DAC modeling series is by no means complete it provides a constructive example of the need to expand the paradigms used in the study of mind, brain, and behavior as opposed to arguing for a shift towards the study of a mindless body.

Acknowledgments

The author is indebted to Peter König for helpful discussions of the manuscript. Part of this project was supported by Swiss National Science Foundation-SPP.

References

[1] N. A., *Unified Theories of Cognition.* Cambridge, Ma.: Harvard University Press, 1990.

[2] A. Newell, "Physical symbol systems," *Cognitive Science*, vol. 4, pp. 135–183, 1980.

[3] A. Newell, J. C. Shaw, and H. A. Simon, "A general problem solving program for a computer," *Computers and Automation*, vol. 8, pp. 10–16, 1959.

[4] R. Vinkhuyzen and P. Verschure, "The legacy of Allen Newell: Unified theories of cognition," *American Journal of Psychology*, vol. 107(3), pp. 454–464, 1994.

[5] J. Fodor, *The Language of Thought.* Cambridge, Ma.: Hardvard University Press, 1975.

[6] S. Pinker, *The Language Instinct.* New York: Morrow, 1994.

[7] J. Fodor, "Fixation of believe and concept acquisition," in *Langauge and Learning: The debate between Jean Piaget and Noam Chomsky* (M. Piatelli-Palmarini, ed.), pp. 143–149, Cambridge, Ma.: Harvard University Press, 1980.

[8] J. McCarthy and P. J. Hayes, "Some philosophical problems from the standpoint of artificial intelligence," *Machine Intelligence*, vol. 4, pp. 463–502, 1969.

[9] J. Searle, "Minds, brains and programs," *Behavioral and Brain Sciences*, vol. 3, pp. 417–424, 1982.

[10] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

[11] W. J. Clancey, "The frame of reference problem in cognitive modeling," in *Proceedings of the Annual Conference of the Cognitive Science Society*, pp. 107–114, Hillsdale, N.J.: Erlbaum, 1989.

[12] L. A. Suchman, *Plans and Situated Actions.* Cambridge: Cambridge University Press, 1987.

[13] P. Verschure, "Formal minds and biological brains," *IEEE expert*, vol. 8, no. 5, pp. 66–75, 1993.

[14] P. Verschure, "Connectionist explanation: Taking positions in the mind-brain dilemma," in *Neural Networks and a New Artificial Intelligence* (G. Dorffner, ed.), pp. 133–188, London: Thompson, 1997.

[15] P. Verschure, "Minds, brains, and robots: Explorations in distributed adaptive control," in *Proceedings of the Second Brazilian-International Conference on Cognitive Science* (A. Soares, ed.), 1996.

[16] D. Rumelhart, J. McClelland, and the PDP research group, *Parallel Distributed Processing: Explorations in the microstructure of cognition.* Cambridge, Ma.: MIT Press, 1986.

[17] R. Brooks, "Intelligence without representation," *Artificial Intelligence*, vol. 47, pp. 139–159, 1991.

[18] R. Pfeifer, "Cognition: Perspectives from autonomous agents," *Robotics and Autonomous Systems*, vol. 15, pp. 47–70, 1995.

[19] P. 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*, pp. 653–658, Hillsdale, N.J.: Erlbaum, 1992.

[20] P. Verschure and A. C. C. Coolen, "Adaptive fields: Distributed representations of classically conditioned associations," *Network*, vol. 2, pp. 189–206, 1991.

[21] P. Verschure, B. Kröse, and R. Pfeifer, "Distributed adaptive control: The self-organization of structured behavior," *Robotics and Autonomous Systems*, vol. 9, pp. 181–196, 1992.

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

[23] P. Verschure and T. Voegtlin, "A bottom-up approach towards the acquisition, retention, and expression of sequential representations: Distributed adaptive control iii," *Neural Networks*, In Press.

[24] L. Kaelbling, M. Littman, and A. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

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

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

[27] P. Verschure and T. Voegtlin, "Convergence properties of predictive hebbian learning," In Preparation.

[28] F. Mondada and P. Verschure, "Modeling system-environment interaction: The complementary roles of simulations and real world artifacts," in *Proceedings of the Second European Conference on Artificial Life* (, ed.), pp. 808–817, Cambridge, Ma.: MIT press, 1993.

[29] P. Verschure, "The cognitive development of an autonomous behaving artifact: the self-organization of categorization, sequencing, and chunking," in *Proceedings of Prerational Intelligence* (H. Cruze, H. Ritter, and J. Dean, eds.), pp. 95–117, Bielefeld: ZiF, 1993.

[30] P. Verschure and T. Voegtlin, "The analysis of different modes of behavioral control: Implications for the study of behavior," In Preparation.

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