A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents

David Vernon, *Senior Member, IEEE,* Giorgio Metta, and Giulio Sandini

***Abstract*— This survey presents an overview of the autonomous development of mental capabilities in computational agents. It does so based on a characterization of cognitive systems as systems which exhibit adaptive, anticipatory, and purposive goal-directed behaviour. We present a broad survey of the various paradigms of cognition, addressing cognitivist (physical symbol systems) approaches, emergent systems approaches, encompassing connectionist, dynamical, and enactive systems, and also efforts to combine the two in hybrid systems. We then review several cognitive architectures drawn from these paradigms. In each of these areas, we highlight the implications and attendant problems of adopting a developmental approach, both from phylogenetic and ontogenetic points of view. We conclude with a summary of the key architectural features that systems capable of autonomous development of mental capabilities should exhibit.**

I. Introduction

THE science and engineering of artificial systems that exhibit mental capabilities has a long history, stretching back over sixty years. The term *mental* is not meant to imply any dualism of mind and body; we use the term in the sense of the complement of physical to distinguish mental development from physical growth. As such, mental faculties entail all aspects of robust behaviour, including perception, action, deliberation, and motivation. As we will see, the term cognition is often used in a similar manner [1].

Cognition implies an ability to understand how things might possibly be, not just now but at some future time, and to take this into consideration when determining how to act. Remembering what happened at some point in the past helps in anticipating future events, so memory is important: using the past to predict the future [2] and then assimilating what does actually happen to adapt and improve the system's anticipatory ability in a virtuous cycle that is embedded in an on-going process of action and perception. Cognition breaks free of the present in a way that allows the system to act effectively, to adapt, and to improve.

But what makes an action the right one to choose? What type of behaviour does cognition enable? These questions open up another dimension of the problem: what motivates cognition? How is perception guided? How are actions selected? And what makes cognition possible? Cognitive skills can improve, but what do you need to get started? What drives the developmental process? In other words, in addition to autonomous perception, action, anticipation, assimilation, and adaptation, there are the underlying motivations to consider. These motivations drive perceptual attention, action selection, and system development, resulting in the long-term robust behaviour we seek from such systems.

From this perspective, a cognitive system exhibits effective behaviour through perception, action, deliberation, communication, and through either individual or social interaction with the environment. The hallmark of a cognitive system is that it can function effectively in circumstances that were not planned for explicitly when the system was designed. That is, it has some degree of plasticity and is resilient in the face of the unexpected [3].

Some authors in discussing cognitive systems go even further. For example, Brachman [4] defines a cognitive computer system as one which — in addition to being able to reason, to learn from experience, to improve its performance with time, and to respond intelligently to things it's never encountered before — would also be able to explain what it is doing and why it is doing it. This would enable it to identify potential problems in following a given approach to carrying out a task or to know when it needed new information in order to complete it. Hollnagel [5] suggests that a cognitive system is able to view a problem in more than one way and to use knowledge about itself and the environment so that it is able to plan and modify its actions on the basis of that knowledge. Thus, for some, cognition also entails a sense of self-reflection in addition to the characteristics of adaptation and anticipation.

Cognition then can be viewed as the process by which the system achieves robust adaptive, anticipatory, autonomous behaviour, entailing embodied perception and action. This viewpoint contrasts with those who see cognition as a distinct component or sub-system of the brain — a module of mind — concerned with rational planning and reasoning, acting on the representations produced by the perceptual apparatus and 'deciding' what action(s) should be performed next. The adaptive, anticipatory, autonomous viewpoint reflects the position of Freeman and Núñez who, in their book *Reclaiming Cognition* [6], re-assert the primacy of action, intention, and emotion in cognition. In the past, as we will see, cognition has been viewed by many as disembodied in principle and a symbol-processing adjunct of perception and action in

Manuscript received January 20, 2006; revised June 1, 2006. This work was supported by the European Commission, Project IST-004370 RobotCub, under Strategic Objective 2.3.2.4: Cognitive Systems.

D. Vernon is with the LIRA-Lab, DIST, University of Genoa, Italy. G. Sandini is with the Italian Institute of Technology (IIT). G. Metta is affiliated with both institutes.

practice. However, this is changing and even proponents of these early approaches now see a much tighter relationship between perception, action, and cognition. For example, consider Anderson *et al.* who say that "There is reason to suppose that the nature of cognition is strongly determined by the perceptual-motor systems, as the proponents of embodied and situated cognition have argued" [7], and Langley who states that "mental states are always grounded in real or imagined physical states, and problem-space operators always expand to primitive skills with executable actions" [8]. Our goal in this paper is to survey the full spectrum of approaches to the creation of artificial cognitive systems with a particular focus on embodied developmental agents.

We begin with a review of the various paradigms of cognition, highlighting their differences and common ground. We then review several cognitive architectures drawn from these paradigms and present a comparative analysis in terms of the key characteristics of embodiment, perception, action, anticipation, adaptation, motiviation, and autonomy. We identify several core considerations shared by contemporary approaches of all paradigms of cognition. We conclude with a summary of the key features that systems capable of autonomous development of mental capabilities should exhibit.

II. The Different Paradigms of Cognition

There are many positions on cognition, each taking a significantly different stance on the nature of cognition, what a cognitive system does, and how a cognitive system should be analyzed and synthesized. Among these, however, we can discern two broad classes: the *cognitivist* approach based on symbolic information processing representational systems, and the *emergent systems* approach, embracing connectionist systems, dynamical systems, and enactive systems, all based to a lesser or greater extent on principles of self-organization [9], [10].

Cognitivist approaches correspond to the classical and still common view that 'cognition is a type of computation' defined on symbolic representations, and that cognitive systems 'instantiate such representations physically as cognitive codes and . . . their behaviour is a causal consequence of operations carried out on these codes' [11]. Connectionist, dynamical, and enactive systems, grouped together under the general heading of emergent systems, argue against the information processing view, a view that sees cognition as 'symbolic, rational, encapsulated, structured, and algorithmic', and argue in favour of a position that treats cognition as emergent, self-organizing, and dynamical [12], [13].

As we will see, the emphasis of the cognitivist and emergent positions differ deeply and fundamentally, and go far beyond a simple distinction based on symbol manipulation. Without wishing to preempt what is to follow, we can contrast the cognitivist and emergent paradigms on twelve distinct grounds: computational operation, representational framework, semantic grounding, temporal constraints, inter-agent epistemology, embodiment, perception, action, anticipation, adaptation, motivation, and autonomy.[1] Let us look briefly at each of these in turn.

The Cognitivist *vs.* Emergent Paradigms of Cognition		
Characteristic	**Cognitivist**	**Emergent**
Computational Operation	Syntactic manipulation of symbols	Concurrent self-organization of a network
Representational Framework	Patterns of symbol tokens	Global system states
Semantic Grounding	Percept-symbol association	Skill construction
Temporal Constraints	Not entrained	Synchronous real-time entrainment
Inter-agent epistemology	Agent-independent	Agent-dependent
Embodiment	Not implied	Cognition implies embodiment
Perception	Abstract symbolic representations	Response to perturbation
Action	Causal consequence of symbol manipulation	Perturbation of the environment by the system
Anticipation	Procedural or probabilistic reasoning typically using *a priori* models	Self-effected traverse of perception-action state space
Adaptation	Learn new knowledge	Develop new dynamics
Motivation	Resolve impasse	Increase space of interaction
Relevance of Autonomy	Not necessarily implied	Cognition implies autonomy

TABLE I

A COMPARISON OF COGNITIVIST AND EMERGENT PARADIGMS OF COGNITION; REFER TO THE TEXT FOR A FULL EXPLANATION.

Computational Operation. Cognitivist systems use rule-based manipulation (*i.e.* syntactic processing) of symbol tokens, typically but not necessarily in a sequential manner. Emergent systems exploit processes of self-organization, self-production, self-maintenance, and self-development, through the concurrent interaction of a network of distributed interacting components.

Representational Framework. Cognitivist systems use patterns of symbol tokens that refer to events in the external world. These are typically the descriptive[2] product of a human designer, usually, but not necessarily, punctate and local. Emergent systems representations are global system states encoded in the dynamic organization of the system's distributed network of components.

Semantic Grounding. Cognitivist systems symbolic representations are grounded through percept-symbol identication by either the designer or by learned association. These representations are accessible to direct human interpretation. Emergent systems ground representations by autonomy-preserving anticipatory and adaptive skill construction. These representations only have meaning insofar as they contribute to the continued viability of the system and are inaccessible to direct human interpretation.

Temporal Constraints. Cognitivist systems are not necessarily entrained by the events in the external world. Emergent systems are entrained and operate synchronously in real-time with events in its environment.

Inter-agent Epistemology. For cognitivist systems an absolute shared epistemology between agents is guaranteed by

[1]There are many possible definitions of autonomy, ranging from the ability of a system to contribute to its own persistence [14] through to the self-maintaining organizational characteristic of living creatures —dissipative far-from equilibrium systems — that enables them to use their own capacities to manage their interactions with the world, and with themselves, in order to remain viable [15].

[2]Descriptive in the sense that the designer is a third-party observer of the relationship between a cognitive system and its environment so that the representational framework is how the designer sees the relationship.

virtue of their positivist view of reality: each agent is embedded in an environment, the structure and semantics of which are independent of the system's cognition. The epistemology of emergent systems is the subjective outcome of a history of shared consensual experiences among phylogentically-compatible agents.

Embodiment. Cognitivist systems do not need to be embodied, in principle, by virtue of their roots in functionalism (which states that cognition is independent of the physical platform in which it is implemented [6]). Emergent systems are intrinsically embodied and the physical instantiation plays a direct constitutive role in the cognitive process [3], [16], [17].

Perception. In cognitivist systems perception provides an interface between the external world and the symbolic representation of that world. Perception abstracts faithful spatio-temporal representations of the external world from sensory data. In emergent systems perception is a change in system state in response to environmental perturbations in order to maintain stability.

Action. In cognitivist systems actions are causal consequences of symbolic processing of internal representations. In emergent systems actions are perturbations of the environment by the system.

Anticipation. In cognitivist systems anticipation typically takes the form of planning using some form of procedural or probabilistic reasoning with some *a priori* model. Anticipation in the emergent paradigm requires the system to visit a number of states in its self-constructed perception-action state space without commiting to the associated actions.

Adaptation. For cognitivism, adaptation usually implies the acquisition of new knowledge whereas in emergent systems, it entails a structural alteration or re-organization to effect a new set of dynamics [95].

Motivation. Motivations, which impinge on perception (through attention), action (through action selection), and adaptation (through the factors that govern change), such as resolving an impasse in a cognitivist system or enlarging the space of interaction in an emergent system [173], [174].

Relevance of Autonomy. Autonomy is not necessarily implied by the cognitivist paradigm whereas it is crucial in the emergent paradigm since cognition is the process whereby an autonomous system becomes viable and effective.

Table I summarizes these points very briefly. The sections that follow discuss the cognitivist and emergent paradigms, as well as hybrid approaches, and draw out each of these issues in more depth.

A. Cognitivist Models

1) An Overview of Cognitivist Models: Cognitive science has its origins in cybernetics (1943-53) in the first efforts to formalize what had up to that point been metaphysical treatments of cognition [9]. The intention of the early cyberneticians was to create a science of mind, based on logic. Examples of progenitors include McCulloch and Pitts and their seminal paper 'A logical calculus immanent in nervous activity' [18]. This initial wave in the development of a science of cognition was followed in 1956 by the development of an approach referred to as *cognitivism*. Cognitivism asserts that cognition involves computations defined over internal representations *qua* knowledge, in a process whereby information about the world is abstracted by perception, and represented using some appropriate symbolic data-structure, reasoned about, and then used to plan and act in the world. The approach has also been labelled by many as the *information processing* (or symbol manipulation) approach to cognition [9], [12], [13], [19]–[23]

Cognitivism has undoubtedly been the predominant approach to cognition to date and is still prevalent. The discipline of cognitive science is often identified with this particular approach [6], [13]. However, as we will see, it is by no means the only paradigm in cognitive science and there are indications that the discipline is migrating away from its stronger interpretations [10].

For cognitivist systems, cognition is representational in a strong and particular sense: it entails the manipulation of explicit symbolic representations of the state and behaviour of the external world to facilitate appropriate, adaptive, anticipatory, and effective interaction, and the storage of the knowledge gained from this experience to reason even more effectively in the future [5]. Perception is concerned with the abstraction of faithful spatio-temporal representations of the external world from sensory data. Reasoning itself is symbolic: a procedural process whereby explicit representations of an external world are manipulated to infer likely changes in the configuration of the world (and attendant perception of that altered configuration) arising from causal actions.

In most cognitivist approaches concerned with the creation of artificial cognitive systems, the symbolic representations (or representational frameworks, in the case of systems that are capable of learning) are the descriptive product of a human designer. This is significant because it means that they can be directly accessed and understood or interpreted by humans and that semantic knowledge can be embedded directly into and extracted directly from the system. However, it has been argued that this is also the key limiting factor of cognitivist systems: these programmer-dependent representations effectively bias the system (or 'blind' the system [24]) and constrain it to an idealized description that is dependent on and a consequence of the cognitive requirements of human activity. This approach works as long as the system doesn't have to stray too far from the conditions under which these descriptions were formulated. The further one does stray, the larger the 'semantic gap' [25] between perception and possible interpretation, a gap that is normally plugged by the embedding of (even more) programmer knowledge or the enforcement of expectation-driven constraints [26] to render a system practicable in a given space of problems.

Cognitivism makes the positivist assumption that 'the world we perceive is isomorphic with our perceptions of it as a geometric environment' [27]. The goal of cognition, for a cognitivist, is to reason symbolically about these representations in order to effect the required adaptive, anticipatory, goal-directed, behaviour. Typically, this approach to cognition will deploy an arsenal of techniques including machine learning,

probabilistic modelling, and other techniques in an attempt to deal with the inherently uncertain, time-varying, and incomplete nature of the sensory data that is being used to drive this representational framework. However, this doesn't alter the fact that the representational structure is still predicated on the descriptions of the designers. The significance of this will become apparent in later sections.

2) Cognitivism and Artificial Intelligence: Since cognitivism and artificial intelligence research have very strong links,[3] it is worth spending some time considering the relationship between cognitivist approaches and classical artificial intelligence, specifically the Newell's and Simon's 'Physical Symbol System' approach to artificial intelligence [20] which has been extraordinarily influential in shaping how we think about intelligence, both natural and computational.

In Newell's and Simon's 1976 paper, two hypotheses are presented:

1) *The Physical Symbol System Hypothesis:* A physical symbol system has the necessary and sufficient means for general intelligent action.
2) *Heuristic Search Hypothesis*. The solutions to problems are represented as symbol structures. A physical-symbol system exercises its intelligence in problem-solving by search, that is, by generating and progressively modifying symbol structures until it produces a solution structure.

The first hypothesis implies that any system that exhibits general intelligence is a physical symbol system *and* any physical symbol system of sufficient size can be configured somehow ('organized further') to exhibit general intelligence.

The second hypothesis amounts to an assertion that symbol systems solve problems by heuristic search, *i.e.* 'successive generation of potential solution structures' in an effective and efficient manner. 'The task of intelligence, then, is to avert the ever-present threat of the exponential explosion of search'.

A physical symbol system is equivalent to an automatic formal system [21]. It is 'a machine that produces through time an evolving collection of symbol structures.' A symbol is a physical pattern that can occur as a component of another type of entity called an expression (or symbol structure): expressions/symbol structures are arrangements of symbols/tokens. As well as the symbol structures, the system also comprises processes that operate on expressions to produce other expressions: 'processes of creation, modification, reproduction, and destruction'. An expression can *designate* an object and thereby the system can either 'affect the object itself or behave in ways depending on the object', or, if the expression designates a process, then the system *interprets* the expression by carrying out the process (see Figure 1).

In the words of Newell and Simon,

> 'Symbol systems are collections of patterns and processes, the latter being capable of producing, destroying, and modifying the former. The most important properties of patterns is that they can designate objects, processes, or other patterns, and that when they designate processes, they can be interpreted. Interpretation means carrying out the designated process. The two most significant classes of symbol systems with which we are acquainted are human beings and computers.'

[3]Some view AI as the direct descendent of cognitivism: "... the positivist and reductionist study of the mind gained an extraordinary popularity through a relatively recent doctrine called *Cognitivism*, a view that shaped the creation of a new field — *Cognitive Science* — and its most hard core offspring: Artificial Intelligence" (emphasis in the original). [6]

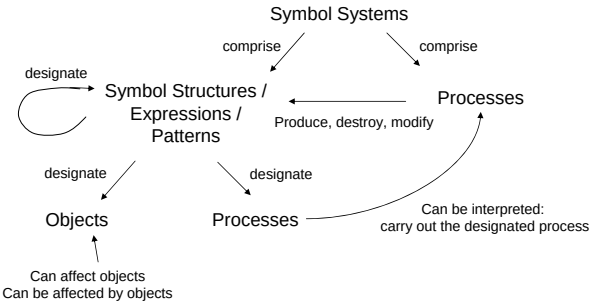

Fig. 1. The essence of a physical symbol system [20].

What is important about this explanation of a symbol system is that it is more general than the usual portrayal of symbol-manipulation systems where symbols designate only objects, in which case we have a system of processes that produces, destroys, and modifies symbols, and no more. Newell's and Simon's original view is more sophisticated. There are two recursive aspects to it: processes can produce processes, and patterns can designate patterns (which, of course, can be processes). These two recursive loops are closely linked. Not only can the system build ever more abstract representations and reason about those representation, but it can modify itself as a function both of its processing, *qua* current state/structure, and of its representations.

Symbol systems can be instantiated and the behaviour of these instantiated systems depend on the details of the symbol system, its symbols, operations, and interpretations, and *not* on the particular form of the instantiation.

The *physical symbol system hypothesis* asserts that a physical symbol system has the necessary and sufficient means for general intelligence. From what we have just said about symbol systems, it follows that intelligent systems, either natural or artificial ones, are effectively equivalent because the instantiation is actually inconsequential, at least in principle.

To a very great extent, cognitivist systems are identically physical symbol systems.

3) Some Cognitivist Systems: Although we will survey cognitivist systems from an architectural point of view in Section III, we mention here a sample of cognitivist systems to provide a preliminary impression of the approach.

The use of explict symbolic knowledge has been used in many cognitivist systems, *e.g.* a cognitive vision system [28] developed for the interpretation of video sequences of traffic behaviour and the generation of a natural language description of the observed environment. It proceeds from signal representations to symbolic representations through several layers

of processing, ultimately representing vehicle behaviour with situation graph trees (SGT). Automatic interpretation of this representation of behaviour is effected by translating the SGT into a logic program (based on fuzzy metric temporal Horn logic). See also [29]–[33] for related work.

The cognitivist assumptions are also reflected well in the model-based approach described in [34], [35] which uses Description Logics, based on First Order Predicate Logic, to represent and reason about high-level concepts such as spatio-temporal object configurations and events.

Probabilistic frameworks have been proposed as an alternative (or sometimes an adjunct [34]) to these types of deterministic reasoning systems. For example, Buxton *et al.* describe a cognitive vision system for interpreting the activities of expert human operators. It exploits dynamic decision networks (DDN) — an extension of Bayesian belief networks to incorporate dynamic dependencies and utility theory [36] — for recognizing and reasoning about activities, and both time delay radial basis function networks (TDRBFN) and hidden markov models (HMM) for recognition of gestures. Although this system does incorporate learning to create the gesture models, the overall symbolic reasoning process, albeit a probabilistic one, still requires the system designer to identify the contextual constraints and their causal dependencies (for the present at least: on-going research is directed at automatically learning the task-based context dependent control strategies) [37]–[39].[4] Recent progress in autonomously constructing and using symbolic models of behaviour from sensory input using inductive logic programming is reported in [40].

The dependence of cognitivist approaches on designer-oriented world-representations is also well exemplified by knowledge-based systems such as those based on ontologies. For example, Maillot *et al.* [41] describe a framework for an ontology-based cognitive vision system which focusses on mapping between domain knowledge and image processing knowledge using a visual concept ontology incorporating spatio-temporal, textural, and colour concepts.

Another architecture for a cognitive vision system is described in [42]. This system comprises a sub-symbolic level, exploiting a viewer-centred $2\frac{1}{2}$D representation based on sensory data, an intermediate pre-linguistic conceptual level based on object-centred 3D superquadric representations, and a linguistic level which uses a symbolic knowledge base. An attentional process links the conceptual and linguistic level.

An adaptable system architecture for observation and interpretation of human activity that dynamically configures its processing to deal with the context in which it is operating is decribed in [43] while a cognitive vision system for autonomous control of cars is described in [44].

Town and Sinclair present a cognitive framework that combines low-level processing (motion estimation, edge tracking, region classification, face detection, shape models, perceptual grouping operators) with high-level processing using a language-based ontology and adaptive Bayesian networks. The system is self-referential in the sense that it maintains an internal representation of its goals and current hypotheses. Visual inference can then be performed by processing sentence structures in this ontological language. It adopts a quintessentially cognitivist symbolic representationalist approach, albeit that it uses probabilistic models, since it requires that a designer identify the "right structural assumptions" and prior probability distributions.

[4]See [36] for a survey of probabilistic generative models for learning and understanding activities in dynamic scenes.

B. Emergent Approaches

Emergent approaches take a very different view of cognition. Here, cognition is the process whereby an autonomous system becomes viable and effective in its environment. It does so through a process of self-organization through which the system is continually re-constituting itself in real-time to maintain its operational identity through moderation of mutual system-environment interaction and co-determination [45]. Co-determination implies that the cognitive agent is specified by its environment and at the same time that the cognitive process determines what is real or meaningful for the agent. In a sense, co-determination means that the agent constructs its reality (its world) as a result of its operation in that world. In this context, cognitive behaviour is sometimes defined as the automatic induction of an ontology: such an ontology will be inherently specific to the embodiment and dependent on the systems history of interactions, *i.e.*, its experiences. Thus, for emergent approaches, perception is concerned with the acquisition of sensory data in order to enable effective action [45] and is dependent on the richness of the action interface [46]. It is not a process whereby the structure of an absolute external environment is abstracted and represented in a more or less isomorphic manner.

Sandini *et al.* have argued that cognition is also the complement of perception [47]. Perception deals with the immediate and cognition deals with longer timeframes. Thus cognition reflects the mechanism by which an agent compensates for the immediate nature of perception and can therefore adapt to and anticipate environmental action that occurs over much longer timescales. That is, cognition is intrinsically linked with the ability of an agent to act prospectively: to operate in the future and deal with what might be, not just what is.

In contrast to the cognitivist approach, many emergent approaches assert that the primary model for cognitive learning is anticipative skill construction rather than knowledge acquisition and that processes that both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems [15]. While cognitivism entails a self-contained abstract model that is disembodied in principle, the physical instantiation of the systems plays no part in the model of cognition [3], [48]. In contrast, emergent approaches are intrinsically embodied and the physical instantiation plays a pivotal role in cognition.

1) Connectionist Models: Connectionist systems rely on parallel processing of non-symbolic distributed activation patterns using statistical properties, rather than logical rules, to process information and achieve effective behaviour [49]. In this sense, the neural network instantiations of the connectionist model 'are dynamical systems which compute functions

that best capture the statistical regularities in training data' [50].

A comprehensive review of connectionism is beyond the scope of this paper. For an overview of the foundation of the field and a selection of seminal papers on connectionism, see Anderson's and Rosenfeld's *Neurocomputing: Foundations of Research* [51] and *Neurocomputing 2: Directions of Research* [52]. Medler provides a succinct survey of the development of connectionism in [49], while Smolensky reviews the field from a mathematical perspective, addressing computational, dynamical, and statistical issues [50], [53]–[55]. Arbib's *Handbook of Brain Theory and Neural Networks* provides very accessible summaries of much of the relevant literature [56].

The roots of connectionism reach back well before the computational era. Although Feldman and Ballard [57] are normally credited with the introduction of the term 'connectionist models' in 1982, the term connectionism has been used as early as 1932 in psychology by Thorndike [58], [59] to signal an expanded form of associationism based, for example, on the connectionist principles clearly evident in William James' model of associative memory,[5] but also anticipating such mechanisms as Hebbian learning. In fact, the introduction to Hebb's book *The Organization of Behaviour* [61], in which he presents an unsupervised neural training algorithm whereby the synaptic strength is increased if both the source and target neurons are active at the same time, contains one of the first usages of the term connectionism [51], p. 43.

We have already noted that cognitivism has some of its roots in earlier work in cognitive science and in McCulloch and Pitts seminal work in particular [18]. McCulloch and Pitts showed that any statement within propositional logic could be represented by a network of simple processing units and, furthermore, that such nets have, in principle, the computational power of a Universal Turing Machine. Depending on how you read this equivalence, McCulloch and Pitts contributed to the foundation of both cognitivism and connectionism.

The connectionist approach was advanced significantly in the late 1950s with the introduction of Rosenblatt's *perceptron* [62] and Selfridge's *Pandemonium* model of learning [63]. Rosenblatt showed that any pattern classification problem expressed in binary notation can be solved by a perceptron network. Although network learning advanced in 1960 with the introduction of the Widrow-Hoff rule, or delta rule, for supervised training in the *Adeline* neural model [64], the problem with perceptron networks was that no learning algorithm existed to allow the adjustment of the weights of the connections between input units and hidden associative units. Consequently, perceptron networks were effectively single-layer networks since learning algorithms could only adjust the connection strength between the hidden units and the output units, the weights governing the connection strength between input and hidden units being fixed by design.

In 1969, Minsky and Papert [65] showed that these perceptrons can only be trained to solve linearly separable problems and couldn't be trained to solve more general problems. As a result, research on neural networks and connectionist models suffered.

With the apparent limitations of perceptions clouding work on network learning, research focussed more on memory and information retrieval and, in particular, on parallel models of associative memory (*e.g.* see [66]). Landmark contributions in this period include McClelland's Interactive Activation and Competition (IAC) model [67] which introduced the idea of competitive pools of mutually-inhibitory neurons and demonstrated the ability of connectionist systems to retrieve specific and general information from stored knowledge about specific instances.

During this period too alternative connectionist models were being put forward in, for example, Grossberg's Adaptive Resonance Theory (ART) [68] and Kohonen's self-organizing maps (SOM) [69], often referred to simply as Kohonen networks. ART, introduced in 1976, has evolved and expanded considerably in the past 30 years to address real-time supervised and unsupervised category learning, pattern classification, and prediction (see [70] for a summary). Kohonen networks produce topological maps in which proximate points in the input space are mapped by an unsupervised self-organizing learning process to an internal network state which preserves this topology: that is, input points (points in pattern space) which are close together are represented in the mapping by points (in weight space) which are close together. Once the unsupervised self-organization is complete, the Kohonen network can be used as either an auto-associative memory or a pattern classifier.

Perceptron-like neural networks underwent a resurgence in the mid 1980s with the development of the parallel distributed processing (PDP) architecture [71] in general and with the introduction by Rumelhart, Hinton, and Williams of the back-propagation algorithm [72], [73]. The back-propagation learning algorithm, also known as the generalized delta rule or GDR as it is a generalization of the Widrow-Hoff delta rule for training Adaline units, overcame the limitation cited by Minsky and Papert by allowing the connections weights between the input units and the hidden units be modified, thereby enabling multi-layer perceptrons to *learn* solutions to problems that are not linearly separable. Although the back-propagation learning rule made its great impact through the work of Rumelhart *et al.*, it had previously been derived independently by Werbos [74], among others [49].

In cognitive science, PDP made a significant contribution to the move away from the sequential view of computational models of mind, towards a view of concurrently-operating networks of mutually-cooperating and competing units, and also in raising an awareness of the importance of the structure of the computing system on the computation.

The standard PDP model represents a static mapping between the input vectors as a consequence of the feed-forward configuration. On the other hand, recurrent networks which have connections that loop back to form circuits, *i.e.* networks in which either the output or the hidden units' activations signals are fed back to the network as inputs, exhibit dynamic

[5]Anderson's and Rosenfeld's collection of seminal papers on neurocomputing [51] opens with Chapter XVI 'Association' from William James' 1890 Psychology, Briefer Course [60].

behaviour.[6] Perhaps the best known type of recurrent network is the Hopfield net [75]. Hopfield nets are fully recurrent networks that act as auto-associative memory[7] or content-addressable memory that can effect pattern completion. Other recurrent networks include Elman nets [76] (with recurrent connections from the hidden to the input units) and Jordan nets [77] (with recurrent connections from the output to the input units). Boltzman machines [78] are variants of Hopfield nets that use stochastic rather than deterministic weight update procedures to avoid problems with the network becoming trapped in local minima during learning.

Multi-layer perceptrons and other PDP connectionist networks typically use monotonic functions, such as hard-limiting threshold functions or sigmoid functions, to activate neurons. The use of non-monotonic activation functions, such as the Gaussian function, can offer computational advantages, *e.g.* faster and more reliable convergence on problems that are not linearly separable.

Radial basis function (RBF) networks [79] also use Gaussian functions but differ from multi-layer perceptrons in that the Gaussian function is used only for the hidden layer, with the input and output layers using linear activation functions.

Connectionist systems continue to have a strong influence on cognitive science, either in a strictly PDP sense such as McClelland's and Rogers' PDP approach to semantic cognition [80]) or in the guise of hybrid systems such as Smolensky's and Legendre's connectionist/symbolic computational architecture for cognition [81], [82].

One of the original motivations for work on emergent systems was disaffection with the sequential, atemporal, and localized character of symbol-manipulation based cognitivism [9]. Emergent systems, on the other hand, depend on parallel, real-time, and distributed architectures. Of itself, however, this shift in emphasis isn't sufficient to constitute a new paradigm and, as we have seen, there are several other pivotal characteristics of emergent systems. Indeed, Freeman and Núñez have argued that more recent systems — what they term neo-cognitivist systems — exploit parallel and distributed computing in the form of artificial neural networks and associative memories but, nonetheless, still adhere to the original cognitivist assumptions [6]. A similar point was made by Van Gelder and Port [83]. We discuss these hybrid systems in Section II-C.

One of the key features of emergent systems, in general, and connectionism, in particular, is that 'the system's connectivity becomes inseparable *from its history of transformations*, and related to the kind of task defined for the system' [9]. Furthermore, symbols play no role.[8] Whereas in the cognitivist approach the symbols are distinct from what they stand for, in the connectionist approach, "meaning relates to the global state of the system" [9]. Indeed, meaning is something attributed by an external third-party observer to the correspondence of a system state with that of the world in which the emergent system is embedded. Meaning is a description attributed by an outside agent: it is not something that is intrinsic to the cognitive system except in the sense that the dynamics of the system reflect the effectiveness of its ability to interact with the world.

Examples of the application of associative learning systems in robotics can be found in [84], [85] where hand-eye coordination is learned by a Kohonen neural network from the association of proprioceptive and exteroceptive stimuli. As well as attempting to model cognitive behaviour, connectionist systems can self-organize to produce feature-analyzing capabilities similar to those of the first few processing stages of the mammalian visual system (*e.g.* centre-surround cells and orientation-selective cells) [86]. An example of a connectionist system which exploits the co-dependency of perception and action in a developmental setting can be found in [87]. This is a biologically-motivated system that learns goal-directed reaching using colour-segmented images derived from a retina-like log-polar sensor camera. The system adopts a developmental approach: beginning with innate inbuilt primitive reflexes, it learns sensorimotor coordination. Radial basis function networks have also been used in cognitive vision systems, for example, to accomplish face detection [38].

2) Dynamical Systems Models: Dynamical systems theory has been used to complement classical approaches in artificial intelligence [88] and it has also been deployed to model natural and artificial cognitive systems [12], [13], [83]. Advocates of the dynamical systems approach to cognition argue that motoric and perceptual systems are both dynamical systems, each of which self-organizes into meta-stable patterns of behaviour.

In general, a dynamical system is an open dissipative non-linear non-equilibrium system: a system in the sense of a large number of interacting components with large number of degrees of freedom, dissipative in the sense that it diffuses energy (its phase space decreases in volume with time implying preferential sub-spaces), non-equilibrium in the sense that it is unable to maintain structure or function without external sources of energy, material, information (and, hence, open). The non-linearity is crucial: as well as providing for complex behaviour, it means that the dissipation is not uniform and that only a small number of the system's degrees of freedom contribute to its behaviour. These are termed *order parameters* (or *collective variables*). Each order parameter defines the evolution of the system, leading to meta-stable states in a multi-stable state space (or phase space). It is this ability to characterize the behaviour of a high-dimensional system with a low-dimensional model that is one of the features that distinguishes dynamical systems from connectionist systems [13].

Certain conditions must prevail before a system qualifies as a cognitive dynamical system. The components of the system must be related and interact with one another: any change in one component or aspect of the system must be dependent on and only on the states of the other components: 'they must be

[6]This recurrent feed-back has nothing to do with the feed-back of error signals by, for example, back-propagation to effect weight adjustment during learning

[7]Hetero-associative memory —or simply associative memory —produces an output vector that is different from the input vector

[8]It would be more accurate to say that symbols should play no role since it has been noted that connectionist systems often fall back in the cognitivist paradigm by treating neural weights as a distributed symbolic representation [83].

interactive and self contained' [83]. As we will see shortly, this is very reminiscent of the requirement for operational closure in enactive systems, the topic of the next section.

Proponents of dynamical systems point to the fact that they provide one directly with many of the characteristics inherent in natural cognitive systems such as multi-stability, adaptability, pattern formation and recognition, intentionality, and learning. These are achieved purely as a function of dynamical laws and consequent self-organization. They require no recourse to symbolic representations, especially those that are the result of human design.

However, Clark [10] has pointed out that the antipathy which proponents of dynamical systems approaches display toward cognitivist approaches rests on rather weak ground insofar as the scenarios they use to support their own case are not ones that require higher level reasoning: they are not 'representation hungry' and, therefore, are not well suited to be used in a general anti-representationalist (or anti-cognitivist) argument. At the same time, Clark also notes that this antipathy is actually less focussed on representations *per se* (dynamical systems readily admit internal states that can be construed as representations) but more on objectivist representations which form an isomorphic symbolic surrogate of an absolute external reality.

It has been argued that dynamical systems allow for the development of higher order cognitive functions, such as intentionality and learning, in a straight-forward manner, at least in principle. For example, intentionality — purposive or goal-directed behaviour — is achieved by the superposition of an intentional potential function on the intrinsic potential function [13]. Similarly, learning is viewed as the modification of already-existing behavioural patterns that take place in a historical context whereby the entire attractor layout (the phase-space configuration) of the dynamical system is modified. Thus, learning changes the whole system as a new attractor is developed.

Although dynamical models can account for several non-trivial behaviours that require the integration of visual stimuli and motoric control, including the perception of affordances, perception of time to contact, and figure-ground bi-stability [13], [89]–[92], the principled feasibility of higher-order cognitive faculties has yet to be validated.

The implications of dynamical models are many: as noted in [12], 'cognition is non-symbolic, nonrepresentational and all mental activity is emergent, situated, historical, and embodied'. It is also socially constructed, meaning that certain levels of cognition emerge from the dynamical interaction between cognitive agents. Furthermore, dynamical cognitive systems are, of necessity, embodied. This requirement arises directly from the fact that the dynamics depend on self-organizing processes whereby the system differentiates itself as a distinct entity through its dynamical configuration and its interactive exploration of the environment.

With emergent systems in general, and dynamical systems in particular, one of the key issues is that cognitive processes are temporal processes that 'unfold' in real-time and synchronously with events in their environment. This strong requirement for synchronous development in the context of its environment again echoes the enactive systems approach set out in the next section. It is significant for two reasons. First, it places a strong limitation on the rate at which the ontogenetic[9] learning of the cognitive system can proceed: it is constrained by the speed of coupling (*i.e.* the interaction) and not by the speed at which internal changes can occur [24]. Natural cognitive systems have a learning cycle measured in weeks, months, and years and, while it might be possible to collapse it into minutes and hours for an artificial system because of increases in the rate of internal adaptation and change, it cannot be reduced below the time-scale of the interaction (or structural coupling; see next section). If the system has to develop a cognitive ability that, *e.g.*, allows it to anticipate or predict action and events that occur over an extended time-scale (*e.g.* hours), it will take at least that length of time to learn. Second, taken together with the requirement for embodiment, we see that the consequent historical and situated nature of the systems means that one cannot short-circuit the ontogenetic development. Specifically, you can't bootstrap an emergent dynamical system into an advanced state of learned behaviour.

With that said, recall from the Introduction that an important characteristic of cognitive systems is their anticipatory capability: their ability to break free of the present. There appears to be a contradiction here. On the one hand, we are saying that emergent cognitive systems are entrained by events in the environment and that their development must proceed in real-time synchronously with the environment, but at the same time that they can break free from this entrainment. In fact, as we will see in Section III, there isn't a contradiction. The synchronous entrainment is associated with the system's interaction with the environment, but the anticipatory capability arises from the internal dynamics of the cognitive system: its capacity for self-organization and self-development involving processes for mirroring and simulating events based on prior experience (brought about historically by the synchronous interaction) but operating internally by self-perturbation and free from the synchronous environmental perturbations of perception and action.

Although dynamical systems theory approaches often differ from connectionist systems on several fronts [12], [13], [83], it is better perhaps to consider them complementary ways of describing cognitive systems, dynamical systems addressing macroscopic behaviour at an emergent level and connectionist systems addressing microscopic behaviour at a mechanistic level [93]. Connectionist systems themselves are, after all, dynamical systems with temporal properties and structures such as attractors, instabilities, and transitions [94]. Typically, however, connectionist systems describe the dynamics in a very high dimensional space of activation potentials and connection strengths whereas dynamical systems theory models describe the dynamics in a low dimensional space where a small number of state variables capture the behaviour of the system as a whole. Schöner argues that this is possible because the macroscopic states of high-dimensional dynamics

[9]Ontogeny is concerned with the development of the system over its lifetime.

and their long-term evolution are captured by the dynamics in that part of the space where instabilities occur: the low-dimensional Center-Manifold [95]. Much of the power of dynamical perspectives comes from this higher-level abstraction of the dynamics [54]. The complementary nature of dynamical systems and connectionist descriptions is emphasized by Schöner and by Kelso [13], [96] who argue that non-linear dynamical systems should be modelled simultaneously at three distinct levels: a boundary constraint level that determines the task or goals (initial conditions, non-specific conditions), a collective variables level which characterize coordinated states, and a component level which forms the realized system (*e.g.* nonlinearly coupled oscillators or neural networks). This is significant because it contrasts strongly with the cognitivist approach, best epitomized by David Marr's advocacy of a three-level hierarchy of abstraction (computational theory, representations and algorithms, and hardware implementation), with modelling at the computational theory level being effected without strong reference to the lower and less abstract levels [97]. This complementary perspective of dynamical systems theory and connectionism enables the investigation of the emergent dynamical properties of connectionist systems in terms of attractors, meta-stability, and state transition, all of which arise from the underlying mechanistic dynamics, and, *vice versa*, it offers the possibility of implementing dynamical systems theory models with connectionist architectures.

3) Enactive Systems Models: Enactive systems take the emergent paradigm even further. In contradistinction to cognitivism, which involves a view of cognition that requires the representation of a given objective pre-determined world [9], [83], enaction [9], [24], [45], [98]–[101] asserts that cognition is a process whereby the issues that are important for the continued existence of a cognitive entity brought out or enacted: co-determined by the entity as it interacts with the environment in which it is embedded. Thus, nothing is 'pre-given', and hence there is no need for symbolic representations. Instead there is an enactive interpretation: a real-time context-based choosing of relevance.

For cognitivism, the role of cognition is to abstract objective structure and meaning through perception and reasoning. For enactive systems, the purpose of cognition is to uncover unspecified regularity and order that can then be construed as meaningful because they facilitate the continuing operation and development of the cognitive system. In adopting this stance, the enactive position challenges the conventional assumption that the world *as the system experiences it* is independent of the cognitive system ('the knower'). Instead, knower and known 'stand in relation to each other as mutual specification: they arise together' [9].

The only condition that is required of an enactive system is *effective action*: that it permit the continued integrity of the system involved. It is essentially a very neutral position, assuming only that there is the basis of order in the environment in which the cognitive system is embedded. From this point of view, cognition is exactly the process by which that order or some aspect of it is uncovered (or constructed) by the system. This immediately allows that there are different forms of reality (or relevance) that are dependent directly on the nature of the dynamics making up the cognitive system. This is not a solipsist position of ungrounded subjectivism, but neither is it the commonly-held position of unique — representable — realism. It is fundamentally a phenomenological position.

The enactive systems research agenda stretches back to the early 1970s in the work of computational biologists Maturana and Varela and has been taken up by others, including some in the main-stream of classical AI [9], [24], [45], [98]–[101].

The goal of enactive systems research is the complete treatment of the nature and emergence of autonomous, cognitive, social systems. It is founded on the concept of autopoiesis – literally *self-production* – whereby a system emerges as a coherent systemic entity, distinct from its environment, as a consequence of processes of self-organization. However, enaction involves different degrees of autopoeisis and three orders of system can be distinguished.

First-order autopoietic systems correspond to cellular entities that achieve a physical identity through structural coupling with their environment. As the system couples with its environment, it interacts with it in the sense that the environmental perturbations trigger structural changes 'that permit it to continue operating'.

Second-order systems are meta-cellular systems that engage in structural coupling with their environment, this time through a nervous system that enables the association of many internal states with the different interactions in which the organism is involved. In addition to processes of self-production, these systems also have processes of self-development. Maturana and Varela use the term operational closure for second-order systems instead of autopoiesis to reflect this increased level of flexibility [45].

Third-order systems exhibit coupling between second-order (*i.e.* cognitive) systems, *i.e.* between distinct cognitive agents. It is significant that second- and third-order systems possess the ability to perturb their own organizational processes and attendant structures. Third-order couplings allow a recurrent (common) ontogenetic drift in which the systems are reciprocally-coupled. The resultant structural adaptation – mutually shared by the coupled systems – gives rise to new phenomonological domains: language and a shared epistemology that reflects (but not abstracts) the common medium in which they are coupled. Such systems are capable of three types of behaviour: (i) the instinctive behaviours that derive from the organizational principles that define it as an autopoietic system (and that emerge from the phylogenetic evolution of the system), (ii) ontogenetic behaviours that derive from the development of the system over its lifetime, and (iii) communicative behaviours that are a result of the third-order structural coupling between members of the society of entities.

The core of the enactive approach is that cognition is a process whereby a system identifies regularities as a consequence of co-determination of the cognitive activities themselves, such that the integrity of the system is preserved. In this approach, the nervous system (and a cognitive agent) does not abstract or 'pick up information' from the environment and therefore the metaphor of calling the brain an information processing device is 'not only ambiguous but patently wrong' [45]. On the contrary, knowledge is the effective use of sensorimotor

contingencies grounded in the structural coupling in which the nervous system exists. Knowledge is particular to the system's history of interaction. If that knowledge is shared among a society of cognitive agents, it is not because of any intrinsic abstract universality, but because of the consensual history of experiences shared between cognitive agents with similar phylogeny and compatible ontogeny.

As with dynamical systems, enactive systems operate in synchronous real-time: cognitive processes must proceed synchronously with events in the systems environment as a direct consequence of the structural coupling and co-determination between system and environment. However, exactly the same point we made about the complementary process of anticipation in dynamical systems applies equally here. And, again, enactive systems are necessarily embodied systems. This is a direct consequence of the requirement of structural coupling of enactive systems. There is no semantic gap in emergent systems (connectionist, dynamical, or enactive): the system builds its own understanding as it develops and cognitive understanding emerges by co-determined exploratory learning. Overall, enactive systems offer a framework by which successively-richer orders of cognitive capability can be achieved, from autonomy of a system through to the emergence of linguistic and communicative behaviours in societies of cognitive agents.

The emergent position in general and the enactive position in particular are supported by recent results which have shown that a biological organism's perception of its body and the dimensionality and geometry of the space in which it is embedded can be deduced (learned or discovered) by the organism from an analysis of the dependencies between motoric commands and consequent sensory data, without any knowledge or reference to an external model of the world or the physical structure of the organism [102], [103]. Thus, the perceived structure of reality could therefore be a consequence of an effort on the part of brains to account for the dependency between their inputs and their outputs in terms of a small number of parameters. Thus, there is in fact no need to rely on the classical idea of an *a priori* model of the external world that is mapped by the sensory apparatus to 'some kind of objective archetype'. The conceptions of space, geometry, and the world that the body distinguishes itself from arises from the sensorimotor interaction of the system, exactly the position advocated in developmental psychology [12]. Furthermore, it is the analysis of the sensory consequences of motor commands that gives rise to these concepts. Significantly, the motor commands are *not* derived as a function of the sensory data. The primary issue is that sensory and motor information are treated simultaneously, and not from either a stimulus perspective or a motor control point of view. As we will see in Section II-C and V-.3, this perception-action co-dependency forms the basis of many artificial cognitive systems.

The enactive approach is mirrored in the work of others. For example, Bickhard [14] introduces the ideas of self-maintenant system and recursive self-maintenant systems. He asserts that

> 'The grounds of cognition are adaptive far-from-equilibrium autonomy — recursively self-maintenant autonomy — not symbol processing nor connectionist input processing. The foundations of cognition are not akin to the computer foundations of program execution, nor to passive connectionist activation vectors.'

Bickhard defines autonomy as the property of a system to contribute to its own persistence. Since there are different grades of contribution, there are therefore different levels of autonomy.

Bickhard introduces a distinction between two types of self-organizing autonomous system:

1) *Self-Maintenant Systems* that make active contributions to their own persistence but do not contribute to the maintenance of the conditions for persistence. Bickhard uses a lighted candle as an example. The flame vapourizes the wax which in turn combusts to form the flame.
2) *Recursive Self-Maintenant Systems* that do contribute actively to the conditions for persistence. These systems can deploy different processes of self-maintenance depending on environmental conditions: "they shift their self-maintenant processes so as to maintain self-maintenance as the environment shifts".

He also distinguishes between two types of stability: (a) *energy well stability* which is equivalent to the stability of systems in thermodynamic equilibrium — no interaction with its environment is required to maintain this equilibrium — and (b) *far from equilibrium stability* which is equivalent to non-thermodynamic equilibrium. Persistence of this state of equilibrium requires that the process or system does not go to thermodynamic equilibrium. These systems are completely dependent for their continued existence on continued contributions of external factors: they require environmental interaction and are necessarily open processes (which nonetheless exhibit closed self-organization).

Self-maintenant and recursive self-maintenant systems are both examples of far-from-equilibrium stability systems.

On the issue of representations in emergent systems, he notes that recursive self-maintenant systems do in fact yield the emergence of representation. Function emerges in self-maintenant systems and representation emerges as a particular type of function ('indications of potential interactions') in recursively self-maintenant systems.

C. Hybrid Models

Considerable effort has also gone into developing approaches which combine aspects of the emergent systems and cognitivist systems [46], [104], [105]. These hybrid approaches have their roots in arguments against the use of explicit programmer-based knowledge in the creation of artificially-intelligent systems [106] and in the development of active 'animate' perceptual systems [107] in which perception-action behaviours become the focus, rather than the perceptual abstraction of representations. Such systems still use representations and representational invariances but it has been argued that these representations should only be constructed by the system itself as it interacts with and explores the world rather than through *a priori* specification or programming so that objects should be represented as 'invariant combinations of percepts and responses where the invariances (which are not

restricted to geometric properties) need to be learned through interaction rather than specified or programmed *a priori*' [46]. Thus, a system's ability to interpret objects and the external world is dependent on its ability to flexibly interact with it and interaction is an organizing mechanism that drives a coherence of association between perception and action. There are two important consequences of this approach of action-dependent perception. First, one cannot have any meaningful direct access to the internal semantic representations, and second cognitive systems must be embodied (at least during the learning phase) [104]. According to Granlund, for instance, action precedes perception and 'cognitive systems need to acquire information about the external world through learning or association' ... 'Ultimately, a key issue is to achieve behavioural plasticity, *i.e.*, the ability of an embodied system to learn to do a task it was not explicitly designed for.' Thus, hybrid systems are in many ways consistent with emergent systems while still exploiting programmer-centred representations (for example, see [108]).

Recent results in building a cognitive vision system on these principles can be found in [109]–[111]. This system architecture combines a neural-network based perception-action component (in which percepts are mediated by actions through exploratory learning) and a symbolic component (based on concepts — invariant descriptions stripped of unnecessary spatial context — can be used in more prospective processing such as planning or communication).

A biologically-motivated system, modelled on brain function and cortical pathways and exploiting optical flow as its primary visual stimulus, has demonstrated the development of object segmentation, recognition, and localization capabilities without any prior knowledge of visual appearance though exploratory reaching and simple manipulation [112]. This hybrid extension of the connectionist system [87] also exhibits the ability to learn a simple object affordance and use it to mimic the actions of another (human) agent.

An alternative hybrid approach, based on subspace learning, is used in [113] to build an embodied robotic system that can achieve appearance-based self-localization using a catadioptric panoramic camera and an incrementally-constructed robust eigenspace model of the environment.

D. Relative Strengths

The foregoing paradigms have their own strengths and weaknesses, their proponents and critics, and they stand at different stages of scientific maturity. The arguments in favour of dynamical systems and enactive systems are compelling but the current capabilities of cognitivist systems are actually more advanced. However, cognitivist systems are also quite brittle.

Several authors have provided detailed critiques of the various approaches. These include, for example, Clark [10], Christensen and Hooker [114], and Crutchfield [115].

Christiansen and Hooker argued [114] that cognitivist systems suffer from three problems: the symbol grounding problem, the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new data),[10] and the combinatorial problem. These problems are one of the reasons why cognitivist models have difficulties in creating systems that exhibit robust sensori-motor interactions in complex, noisy, dynamic environments. They also have difficulties modelling the higher-order cognitive abilities such as generalization, creativity, and learning [114]. According to the Christensen and Hooker, and as we have remarked on several occasions, cognitivist systems are poor at functioning effectively outside narrow, well-defined problem domains.

Enactive and dynamical systems should in theory be much less brittle because they emerge through mutual specification and co-development with the environment, but our ability to build artificial cognitive systems based on these principles is actually very limited at present. To date, dynamical systems theory has provided more of a general modelling framework rather than a model of cognition [114] and has so far been employed more as an analysis tool than as a tool for the design and synthesis of cognitive systems [114], [117]. The extent to which this will change, and the speed with which it will do so, is uncertain. Hybrid approaches appear to some to offer the best of both worlds: the adaptability of emergent systems (because they populate their representational frameworks through learning and experience) but the advanced starting point of cognitivist systems (because the representational invariances and representational frameworks don't have to be learned but are designed in). However, it is unclear how well one can combine what are ultimately highly antagonistic underlying philosophies. Opinion is divided, with arguments both for (*e.g.* [10], [110], [115]) and against (*e.g.* [114]).

A cognitive system is inevitably going to be a complex system and it will exhibit some form of organization, even if it isn't the organization suggested by cognitivist approaches. Dynamical systems theory doesn't, at present, offer much help in identifying this organization since the model is a state-space dynamic which is actually abstracted away from the physical organization of the underlying system [114]. The required organization may not necessarily follow the top-down functional decomposition of AI but some appropriate form of functional organization may well be required. We will return to this issue and discuss it in some depth in Section III on cognitive architectures.

Dynamical systems at present provides more of a general modelling framework rather than a model of cognition is well made and others have made a similar point that dynamical systems approaches has so far been employed more as an analysis tool than as a tool for the design and synthesis of cognitive systems [114], [117].

Clark suggests that one way forward is the development of a form of 'dynamic computationalism' in which dynamical elements form part of an information-processing system [10]. This idea is echoed by Crutchfield [115] who, whilst agreeing that dynamics are certainly involved in cognition, argues that dynamics *per se* are "not a substitute for information processing and computation in cognitive processes" but neither

[10]In the cognitivist paradigm, the frame problem has been expressed in slightly different but essentially equivalent terms: how can one build a program capable of inferring the effects of an action without reasoning explicitly about all its perhaps very many non-effects? [116]

are the two approaches incompatible. He holds that a synthesis of the two can be developed to provide an approach that does allow dynamical state space structures to support computation. He proposes 'computational mechanics' as the way to tackle this synthesis of dynamics and computation. However, this development requires that dynamics itself needs to be extended significantly from one which is deterministic, low-dimensional, and time asymptotic, to one which is stochastic, distributed and high dimensional, and reacts over transient rather than asymptotic time scales. In addition, the identification of computation with digital or discrete computation has to be relaxed to allow for other interpretations of what it is to compute.

III. Cognitive Architectures

Although used freely by proponents of the cognitivist, emergent, and hybrid approaches to cognitive systems, the term cognitive architecture originated with the seminal cognitivist work of Newell *et al.* [118]–[120]. Consequently, the term has a very specific meaning in this paradigm where cognitive architectures represent attempts to create unified theories of cognition [7], [119], [121], *i.e.* theories that cover a broad range of cognitive issues, such as attention, memory, problem solving, decision making, learning, from several aspects including psychology, neuroscience, and computer science. Newell's Soar architecture [120], [122]–[124], Anderson's ACT-R architecture [7], [125], and Minsky's *Society of Mind* [126] are all candidate unified theories of cognition. For emergent approaches to cognition, which a focus on development from a primitive state to a fully cognitive state over the lifetime of the system, the architecture of the system is equivalent to its phylogenetic configuration: the initial state from which it subsequently develops.

In the cognitivist paradigm, the focus in a cognitive architecture is on the aspects of cognition that are constant over time and that are relatively independent of the task [8], [127], [128]. Since cognitive architectures represent the fixed part of cognition, they cannot accomplish anything in their own right and need to be provided with or acquire knowledge to perform any given task. This combination of a given cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model*. In most cognitivist systems the knowledge incorporated into the model is normally determined by the human designer, although there is in increasing use of machine learning to augment and adapt this knowledge. The specification of a cognitive architecture consists of its representational assumptions, the characteristics of its memories, and the processes that operate on those memories. The cognitive architecture defines the manner in which a cognitive agent manages the primitive resources at its disposal [129]. For cognitivist approaches, these resources are the computational system in which the physical symbol system is realized. The architecture specifies the formalisms for knowledge representations and the memory used to store them, the processes that act upon that knowledge, and the learning mechanisms that acquire it. Typically, it also provides a way of programming the system so that intelligent systems can be instantiated in some application domain [8].

For emergent approaches, the need to identify an architecture arises from the intrinsic complexity of a cognitive system and the need to provide some form of structure within which to embed the mechanisms for perception, action, adaptation, anticipation, and motivation that enable the ontogenetic development over the system's life-time. It is this complexity that distinguishes an emergent developmental cognitive architecture from a simple connectionist robot control system that typically learns associations for specific tasks, *e.g.* the Kohonen self-organized net cited in [84]. In a sense, the cognitive architecture of an emergent system corresponds to the innate capabilities that are endowed by the system's phylogeny and which don't have to be learned but of course which may be developed further. These resources facilitate the system's ontogensis. They represent the initial point of departure for the cognitive system and they provide the basis and mechanism for its subsequent autonomous development, a development that may impact directly on the architecture itself. As we have stated already, the autonomy involved in this development is important because it places strong constraints on the manner in which the system's knowledge is acquired and by which its semantics are grounded (typically by autonomy-preserving anticipatory and adaptive skill construction) and by which an inter-agent epistemology is achieved (the subjective outcome of a history of shared consensual experiences among phylogenetically-compatible agents); see Table I.

It is important to emphasize that the presence of innate capabilities in emergent systems does *not* in any way imply that the architecture is functionally modular: that the cognitive system is comprised of distinct modules each one carrying out a specialized cognitive task. If a modularity is present, it may be because it develops this modularity through experience as part of its ontogenesis or epigenesis rather than being prefigured by the phylogeny of the system (*e.g.* see Karmiloff-Smith's theory of representational redescription, [130], [131]). Even more important, it does not necessarily imply that the innate capabilities are hard-wired cognitive skills as suggested by nativist psychology (*e.g.* see Fodor [132] and Pinker [133]).[11] At the same time, neither does it necessarily imply that the cognitive system is a blank slate, devoid of any innate cognitive structures as posited in Piaget's constructivist view of cognitive development [135];[12] at the very least there must exist a mechanism, structure, and organization which allows the cognitive system to be autonomous, to act effectively to some limited extent, and to develop that autonomy.

Finally, since the emergent paradigm sits in opposition to the two pillars of cognitivism — the dualism that posits the logical separation of mind and body, and the functionalism that posits that cognitive mechanisms are independent of the physical platform [6] — it is likely that the architecture will reflect or recognize in some way the morphology of the physical body

[11]More recently, Fodor [134] asserts that modularity applies only to local cognition (*e.g.* recognizing a picture of Mount Whitney) but not global cognition (*e.g.* deciding to trek the John Muir Trail).

[12]Piaget founded the constructivist school of cognitive development whereby knowledge is not implanted *a priori* (*i.e.* phylogenetically) but is discovered and constructed by a child through active maniulation of the environment.

Cognitivist	Emergent	Hybrid
Soar	AAR	HUMANOID
EPIC	Global Workspace	Cerebus
ACT-R	I-C SDAL	Cog: Theory of Mind
ICARUS	SASE	Kismet
ADAPT	DARWIN	

TABLE II
THE COGNITIVE ARCHITECTURES REVIEWED IN THIS SECTION.

of which it is embedded and of which it is an intrinsic part.

Having established these boundary conditions for cognitivist and emergent cognitive architectures (and implicitly for hybrid architectures), for the purposes of this review the term cognitive architecture will the taken in the general and non-specific sense. By this we mean the minimal configuration of a system that is necessary for the system to exhibit cognitive capabilities and behaviours: the specification of the components in a cognitive system, their function, and their organization as a whole. That said, we do place particular emphasis on the need of systems that are developmental and emergent, rather than pre-configured.

Below, we will review several cognitive architectures drawn from the cognitivist, emergent, and hybrid traditions, beginning with some of the best known cognitivist ones. Table II lists the cognitive architectures reviewed under each of these three headings. Following this review, we present a comparative analysis of these architectures using a subset of the twelve paradigm characteristics we discussed in Section II: computational operation, representational framework, semantic grounding, temporal constraints, inter-agent epistemology, role of physical instantiation, perception, action, anticipation, adaptation, motivation, embodiment, autonomy.

A. The Soar Cognitive Architecture

The Soar system [120], [122]–[124] is Newell's candidate for a Unified Theory of Cognition [119]. It is a production (or rule-based) system[13] that operates in a cyclic manner, with a production cycle and a decision cycle. It operates as follows. First, all productions that match the contents of declarative (working) memory fire. A production that fires may alter the state of declarative memory and cause other productions to fire. This continues until no more productions fire. At this point, the decision cycle begins in which a single action from several possible actions is selected. The selection is based on stored action preferences. Thus, for each decision cycle there may have been many production cycles. Productions in Soar are low-level; that is to say, knowledge is encapsulated at a very small grain size.

One important aspect of the decision process concerns a process known as *universal sub-goaling*. Since there is no guarantee that the action preferences will be unambiguous or that they will lead to a unique action or indeed any action, the decision cycle may lead to an 'impasse'. If this happens, Soar sets up an new state in a new problem space — sub-goaling — with the goal of resolving the impasse. Resolving one impasse may cause others and the sub-goaling process continues. It is assumed that degenerate cases can be dealt with (*e.g.* if all else fails, choose randomly between two actions). Whenever an impasse is resolved, Soar creates a new production rule which summarizes the processing that occurred in the sub-state in solving the sub-goal. Thus, resolving an impasse alters the system super-state, *i.e.* the state in which the impasse originally occurred. This change is called a result and becomes the outcome of the production rule. The condition for the production rule to fire is derived from a dependency analysis: finding what declarative memory items matched in the course of determining the result. This change in state is a form of learning and it is the only form that occurs in Soar, *i.e.* Soar only learns new production rules. Since impasses occur often in Soar, learning is pervasive in Soar's operation.

[13]A production is effectively an IF-THEN condition-action pair. A production system is a set of production rules and a computational engine for interpreting or executing productions.

B. EPIC — Executive Process Interactive Control

EPIC [136] is a cognitive architecture that was designed to link high-fidelity models of perception and motor mechanisms with a production system. An EPIC model requires both knowledge encapsulated in production rules and perceptual-motor parameters. There are two types of parameter: standard or system parameters which are fixed for all tasks (*e.g.* the duration of a production cycle in the cognitive processor: 50 ms) and typical parameters which have conventional values but can vary between tasks (*e.g.* the time required to effect recognition of shape by the visual processor: 250 ms).

EPIC comprises a cognitive processor (with a production rule interpreter and a working memory), and auditory processor, a visual processor, an oculo-motor processor, a vocal motor processor, a tactile processor, and an manual motor processor. All processors run in parallel. The perceptual processors simply model the temporal aspects of perception: they don't perform any perceptual processing *per se*. For example, the visual processor doesn't do pattern recognition. Instead, it only models the time it takes for a representation of a given stimulus to be transferred to the declarative (working) memory. A given sensory stimulus may have several possible representations (*e.g.* colour, size, ...) with each representation possibly delivered to the working memory at different times. Similarly, the motor processors are not concerned with the torques required to produce some movement; instead, they are only concerned with the time it takes for some motor output to be produced after the cognitive processor has requested it.

There are two phases to movements: a preparation phase and an execution phase. In the preparation phase, the timing is independent of the number of features that need to be prepared to effect the movement but may vary depending on whether the features have already been prepared in the previous movement. The execution phase is concerned with the timing for the implementation of a movement and, for example, in the case of hand or finger movements the time is governed by Fitt's Law.

Like Soar, the cognitive processor in EPIC is a production system in which multiple rules can fire in one production cycle.

However, the productions in EPIC have a much larger grain size than Soar productions.

Arbitration of resources (*e.g.* when two tasks require a single resource) is handled by 'executive' knowledge: productions which implement executive knowledge do so in parallel with productions for task knowledge.

EPIC does not have any learning mechanism.

C. ACT-R — Adaptive Control of Thought - Rational

The ACT-R [7], [125] cognitive architecture is another approach to creating an unified theory of cognition. It focusses on the modular decomposition of cognition and offers a theory of how these modules are integrated to produce coherent cognition. The architecture comprises five specialized modules, each devoted to processing a different kind of information (see Figure 2). There is a vision module for determining the identity and position of objects in the visual field, a manual module for controlling hands, a declarative module for retrieving information from long-term information, and a goal module for keeping track of the internal state when solving a problem. Finally, it also has a production system that coordinates the operation of the other four modules. It does this indirectly via four buffers into which each module places a limited amount of information.

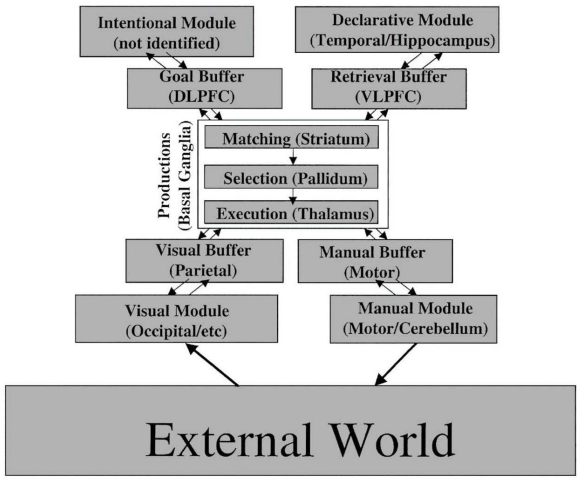

Fig. 2. The ACT-R Cognitive Architecture (from [7]).

ACT-R operates in a cyclic manner in which the patterns of information held in the buffers (and determined by external world and internal modules) are recognized, a single production fires, and the buffers are updated. It is assumed that this cycle takes approximately 50 ms.

There are two serial bottle-necks in ACT-R. One is that the content of any buffer is limited to a single declarative unit of knowledge, called a 'chunk'. This implies that only one memory can be retrieved at a time and indeed that a single object can be encoded in the visual field at any one time. The second bottle-neck is that only one production is selected to fire in any one cycle. This contrasts with both Soar and EPIC both of which allow many productions to fire. When multiple production rules are capable of firing, an arbitration procedure called conflict resolution is activated.

Whilst early incarnations of ACT-R focussed primarily on the production system, the importance of perceptuo-motor processes in determining the nature of cognition is recognized by Anderson *et al.* in more recent versions [7], [121]. That said, the perceptuo-motor system in ACT-R is based on the EPIC architecture [136] which doesn't deal directly with real sensors or motors but simply models the basic timing behaviour of the perceptual and motor systems. In effect, it assumes that the perceptual system has already parsed the visual data into objects and associated sets of features for each object [125]. Anderson *et al.* recognize that this is a shortcoming, remarking that ACT-R implements more a theory of visual attention than a theory of perception, but hope that the ACT-R cognitive architecture will be compatible with more complete models of perceptual and motor systems. The ACT-R visual module differs somewhat from the EPIC visual system in that it is separated into two sub-modules, each with its own buffer, one for object localization and associated with the dorsal pathway, and the other for object recognition and associated with the ventral pathway. Note that this sharp separation of function between the ventral and dorsal pathways has been challenged by recent neurophysiological evidence which points to the interdependence between the two pathways [137], [138]. When the production system requests information from the localization module, it can supply constraints in the form of attribute-value pairs (*e.g.* colour-red) and the localization module will then place a chunk in its buffer with the location of some object that satisfies those constraints. The production system queries the recognition system by placing a chunk with location information in its buffer; this causes the visual system to subsequently place a chunk representing the object at that location in its buffer for subsequent processing by the production system. This is a significant idealization of the perceptual process.

The goal module keeps track of what the intentions of the system architecture (in any given application) so that the behaviour of the system will support the achievement of that goal. In effect, it ensures that the operation of the system is consistent in solving a given problem (in the words of Anderson *et al.* "it maintains local coherence in a problem-solving episode").

On the other hand, the information stored in the declarative memory supports long-term personal and cultural coherence. Together with the production system, which encapsulates procedural knowledge, it forms the core of the ACT-R cognitive system. The information in the declarative memory augments symbolic knowledge with subsymbolic representations in that the behaviour of the declarative memory module is dependent of several numeric parameters: the activation level of a chunk, the probability of retrieval of a chunk, and the latency of retrieval. The activation level is dependent on a learned base level of activation reflecting its overall usefulness in the past, and an associative component reflecting its general usefulness in the current context. This associative component is a weighted sum of the element connected with the current goal. The probability of retrieval is an inverse exponential function

of the activation and a given threshold, while the latency of a chunk that is retrieved (*i.e.* that exceeds the threshold) is an exponential function of the activation.

Procedural memory is encapsulated in the production system which coordinates the overall operation of the architecture. Whilst several productions may qualify to fire, only one production is selected. This selection is called conflict resolution. The production selected is the one with the highest utility, a factor which is a function of an estimate of the probability that the current goal will be achieved if this production is selected, the value of the current goal, and an estimate of the cost of selecting the production (typically proportional to time), both of which are learned in a Bayesian framework from previous experience with that production. In this way, ACT-R can adapt to changing circumstances [121].

Declarative knowledge effectively encodes things in the environment while procedural knowledge encodes observed transformations; complex cognition arises from the interaction of declarative and procedureal knowledge [125]. A central feature of the ACT-R cognitive architecture is that these two types of knowledge are tuned in specific application by encoding the statistics of knowledge. Thus, ACT-R learns sub-symbolic information by adjusting or tuning the knowledge parameters. This sub-symbolic learning distiguishes ACT-R from the symbolic (production-rule) learning of Soar.

Anderson *et al.* suggest that four of these five modules and all four buffers correspond to distinct areas in the human brain. Specifically, the goal buffer corresponds to the dorsolateral pre-frontal cortex (DLPFC), the declarative module to the temporal hippocampus, the retrieval buffer (which acts as the interface between the delarative module and the production system) to the ventrolateral pre-frontal cortex (VLPFC), the visual buffer to the parietal area, the visual module to the occipital area, the manual buffer to the motor system, the manual module to the motor system and cerebellum, the production system to the basal ganglia. The goal module is not associated with a specific brain area. Anderson *et al.* hypothesize that part of the basal ganglia, the striatum, performs a pattern recognition function. Another part, the pallidium, performs a conflict resolution function, and the thalamus controls the execution of the productions.

Like Soar, ACT-R has evolved significantly over several years [125]. It is currently in Version 5.0 [7].

D. The ICARUS Cognitive Architecture

The ICARUS cognitive architecture [8], [139]–[141] follows in the tradition of other cognitivist architectures, such ACT-R, Soar, and EPIC, exploiting symbolic representations of knowledge, the use of pattern matching to select relevant knowledge elements, operation according to the conventional recognize-act cycle, and an incremental approach to learning. In this, ICARUS adheres strictly to the Newell and Simon's physical symbol system hypothesis [20] which states that symbolic processing is a necessary and sufficient condition for intelligent behaviour. However, ICARUS goes further and claims that mental states are always grounded in either real or imagined physical states, and *vice versa* that problem-space symbolic operators always expand to actions that can be effected or executed. Langley refers to this as the *symbolic physical system* hypothesis. This assertion of the importance of action and perception is similar to recent claims by others in the cognitivist community such as Anderson *et al.* [7].

There are also some other important difference between ICARUS and other cognitivist architectures. ICARUS distinguishes between concepts and skills, and devotes two different types of representation and memory for them, with both long-term and short-term variants of each. Conceptual memory encodes knowledge about general classes of objects and relations among them whereas skill memory encodes knowledge about ways to act and achieve goals. ICARUS forces a strong correspondence between short-term and long-term memories, with the latter containing specific instances of the long-term structures. Furthermore, ICARUS adopts a strongly hierarchical organization for its long-term memory, with conceptual memory directing bottom-up inference and skill memory structuring top-down selection of actions.

Langley notes that incremental learning is central to most cognitivist cognitive architectures, in which new cognitive structures are created by problem solving when an impasse is encountered. ICARUS adopts a similar stance so that when an execution module cannot find an applicable skill that is relevant to the current goal, it resolves the impasse by backward chaining.

E. ADAPT — A Cognitive Architecture for Robotics

Some authors, e.g. Benjamin *et al.* [142], argue that existing cognitivist cognitive architectures such as Soar, ACT-R, and EPIC, don't easily support certain mainstream robotics paradigms such as adaptive dynamics and active perception. Many robot programs comprise several concurrent distributed communicating real-time behaviours and consequently these architectures are not suited since their focus is primarily on "sequential search and selection", their learning mechanisms focus on composing sequential rather than concurrent actions, and they tend to be hierarchically-organized rather than distributed. Benjamin *et al.* don't suggest that you cannot address such issues with these architectures but that they are not central features. They present a different cognitive architecture, ADAPT — Adaptive Dynamics and Active Perception for Thought, which is based on Soar but also adopts features from ACT-R (such as long-term declarative memory in which sensori-motor schemas to control perception and action are stored) and EPIC (all the perceptual processes fire in parallel) but the low-level sensory data is placed in short-term working memory where it is processed by the cognitive mechanism. ADAPT has two types of goals: task goals (such as 'find the blue object') and architecture goals (such as 'start a schema to scan the scene'). It also has two types of actions: task actions (such as 'pick up the blue object') and architectural actions (such as 'initiate a grasp schema'). While the architectural part is restricted to allow only one goal or action at any one time, the task part has no such restrictions and many task goals and actions — schemas — can be operational at the same time. The architectural goals and actions are represented procedurally (with productions) while the task

goals and actions are represented declaratively in working memory as well as procedurally.

F. Autonomous Agent Robotics

Autonomous agent robotics (AAR) and behaviour-based systems represents an emergent alternative to cognitivist approaches. Instead of a cognitive system architecture that is based on a decomposition into functional components (*e.g.* representation, concept formation, reasoning), an AAR architecture is based on interacting *whole* systems. Beginning with simple whole systems that can act effectively in simple circumstances, layers of more sophisticated systems are added incrementally, each layer subsuming the layers beneath it. This is the subsumption architecture introduced by Brooks [143]. Christensen and Hooker [114] argue that AAR is not sufficient either as a principled foundation for a general theory of situated cognition. One limitation includes the explosion of systems states that results from the incremental integration of sub-systems and the consequent difficulty in coming up with an initial well-tuned design to produce coordinated activity. This in turn imposed a need from some form of self-management, something not included in the scope of the original subsumption architecture. A second limitation is that it becomes increasingly problematic to rely on environmental cues to achieve the right sequence of actions or activities as the complexity of the task rises. AAR is also insufficient for the creation of a comprehensive theory of cognition: as the subsumption architecture can't be scaled to provide higher-order cognitive faculties (it can't explain self-directed behaviour) and even though the behaviour of an AAR system may be very complex it is still ultimately a reactive system.

Christensen and Hooker note that Brooks has identified a number of design principles to deal with these problems. These include motivation, action selection, self-adaption, and development. Motivation provides context-sensitive selection of preferred actions, while coherence enforces an element of consistency in chosen actions. Self-adaption effects continuous self-calibration among the sub-systems in the subsumption architecture, while development offers the possibility of incremental open-ended learning.

We see here a complementary set of self-management processes, signalling the addition of system-initiated contributions to the overall interaction process and complementing the environmental contributions that are typical of normal subsumption architectures. It is worth remarking that this quantum jump in complexity and organization is reminiscent of the transition from level one autopoietic systems to level two, where the central nervous system then plays a role in allowing the system to perturb itself (in addition to the environmental perturbations of a level 1 system).

G. A Global Workspace Cognitive Architecture

Shanahan [116], [144]–[146] proposes a biologically-plausible brain-inspired neural-level cognitive architecture in which cognitive functions such as anticipation and planning are realized through internal simulation of interaction with the environment. Action selection, both actual and internally simulated, is mediated by affect. The architecture is based on an external sensori-motor loop and an internal sensori-motor loop in which information passes though multiple competing cortical areas and a global workspace.

In contract to manipulating declarative symbolic representations as cognitivist architectures do, cognitive function is achieved here through topographically-organized neural maps which can be viewed as a form of analogical or iconic representation whose structure is similar to the sensory input of the system whose actions they mediate.

Shanahan notes that such analogical representations are particularly appropriate in spatial cognition which is a crucial cognitive capacity but which is notoriously difficult with traditional logic-based approaches. He argues that the semantic gap between sensory input and analogical representations is much smaller than with symbolic language-like representations

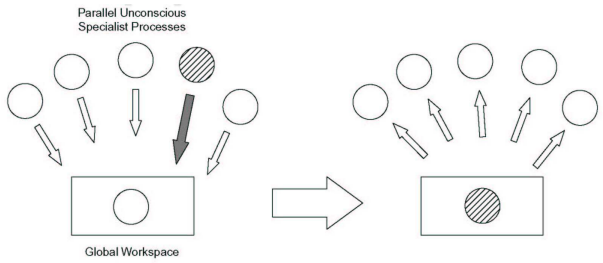

Fig. 3. The Global Workspace Theory cognitive architecture: 'winner-take-all' coordination of competing concurrent processes (from [144]).

Shanahan's cognitive architecture is founded also upon the fundamental importance of parallelism as a constituent component in the cognitive process as opposed to being a mere implementation issue. He deploys the *global workspace* model [147], [148] of information flow in which a sequence of states emerges from the interaction of many separate parallel processes (see Figure 3). These specialist processes compete and co-operate for access to a global workspace. The winner(s) of the competition gain(s) controlling access to the global access and can then broadcast information back to the competing specialist processes. Shanahan argues that this type of architecture provides an elegant solution to the frame problem.

Shanahan's cognitive architecture is comprised of the following components: a first-order sensori-motor loop, closed externally through the world, and a higher-order sensori-motor loop, closed internally through associative memories (see Figure 3). The first-order loop comprises the sensory cortex and the basal ganglia (controlling the motor cortex), together providing a reactive action-selection sub-system. The second-order loop comprises two associative cortex elements which carry out off-line simulations of the system's sensory and motor behaviour, respectively. The first associative cortex simulates a motor output while the second simulates the sensory stimulus expected to follow from a given motor output.

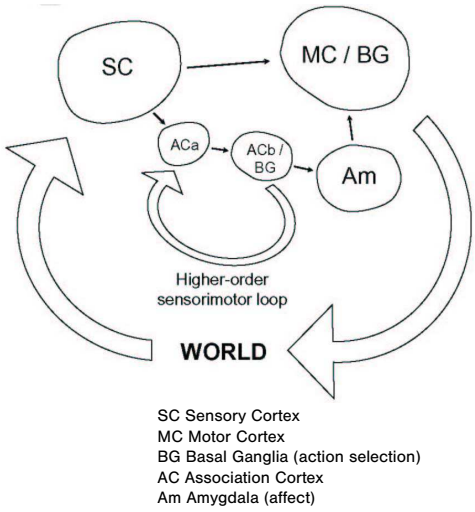

Fig. 4. The Global Workspace Theory cognitive architecture: achieving prospection by sensori-motor simulation (from [144]).

The higher-order loop effectively modulates basal ganglia action selection in the first-order loop via an affect-driven amygdala component. Thus, this cognitive architecture is able to anticipate and plan for potential behaviour through the exercise of its "imagination" (*i.e.* its associative internal sensori-motor simulation. The global workspace doesn't correspond to any particular localized cortical area. Rather, it is a global communications network.

The architecture is implemented as a connectionist system using G-RAMs: generalized random access memories [149]. Interpreting its operation in a dynamical framework, the global workspace and competing cortical assemblies each define an attractor landscape. The perceptual categories constitute attactors in a state space that reflects the structure of the raw sensory data. Prediction is achieved by allowing the higher-order sensori-motor loop to traverse along a simulated trajectory in that state space so that the global workspace visits a sequence of attractors. The system has been validated in a Webot [150] simulation environment.

H. Self-Directed Anticipative Learning

Christensen and Hooker propose a new emergent interactivist-constructivist (I-C) approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [15]. This approach falls under the broad heading of dynamical embodied approaches in the non-cognitivist paradigm. They assert first the primary model for cognitive learning is anticipative skill construction and that processes that both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems. For them, intelligence is a continuous management process that has to support the need to achieve autonomy in a living agent, distributed dynamical organization, and the need to produce functionally coherent activity complexes that match the constraints of autonomy with the appropriate organization of the environment across space and time through interaction. In presenting their approach they use the term "explicit norm signals" for the signals that a system uses to differentiate an appropriate context performing an action. These norm signals reflect conditions for the (maintenance) of the system's autonomy (*e.g.* hunger signals depleted nutritional levels). The complete set of norm signals is termed the norm matrix. They then distinguish between two levels of management: low-order and high-order. Low-order management employs norm signals which differentiate only a narrow band of the overall interaction process of the system (*e.g.* a mosquito uses heat tracking and CO_2 gradient tracking to seek blood hosts). Since it uses only a small number of parameters to direct action, success ultimately depends on simple regularity in the environment. These parameters also tend to be localized in time and space. On the other hand, high-order management strategies still depend to an extent on regularity in the environment but exploit parameters that are more extended in time and space and use more aspects of the interactive process, including the capacity to anticipate and evaluate the system's performance, to produce effective action (and improve performance). This is the essence of self-directedness. "Self-directed systems anticipate and evaluate the interaction process and modulate system action accordingly". The major features of self-directedness are action modulation ("generating the right kind of extended interaction sequences"), anticipation ("who will/should the interaction go?", evaluation ("how did the evaluation go?"), and constructive gradient tracking ("learning to improve performance").

I. A Self-Affecting Self-Aware (SASE) Cognitive Architecture

Weng [151]–[153] introduced an emergent cognitive architecture that is specifically focussed on the issue of development by which he means that the processing accomplished by the architecture is not specified (or programmed) *a priori* but is the result of the real-time interaction of the system with the environment including humans. Thus, the architecture is not specific to tasks, which are unknown when the architecture is created or programmed, but is capable of adapting and developing to learn both the tasks required of it and the manner in which to achieve the tasks.

Weng refers to his architecture as a Self-Aware Self-Effecting (SASE) system (see Figure 5). The architecture entails an important distinction between the sensors and effectors that are associated with the environment (including the system's body and thereby including proprioceptive sensing) and those that are associated with the system's 'brain' or central nervous system (CNS). Only those systems that have explicit mechanisms for sensing and affecting the CNS qualify as SASE architectures. The implications for development are significant: the SASE architecture is configured with no knowledge of the tasks it will ultimately have to perform, its brain or CNS are not directly accessible to the (human) designers once it is launched, and after that the only way a human can affect the agent is through the external sensors and effectors. Thus, the SASE architecture is very faithful to

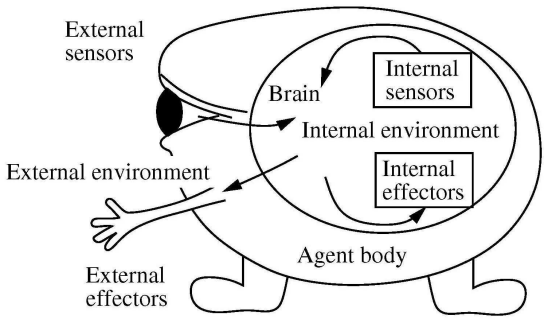

Fig. 5. The Self-Aware Self-Effecting (SASE) architecture (from [153]).

the emergent paradigms of cognition, especially the enactive approach: its phylogeny is fixed and it is only through ontogenetic development that the system can learn to operate effectively in its environment.

The concept of self-aware self-effecting operation is similar to the level 2 autopoietic organizational principles introduced by Matura and Varela [45] (*i.e.* both self-production and self-development) and is reminiscent of the recursive self-maintenant systems principles of Bickhard [14] and Christensen's and Hooker's interactivist-constructivist approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [15]. Weng's contribution differs in that he provides a specific computational framework in which to implement the architecture. Weng's cognitive architecture is based on Markov Decision Processes (MDP), specifically a developmental observation-driven self-aware self-effecting Markov Decision Process (DOSASE MDP). Weng places this particular architecture in a spectrum of MDPs of varying degrees of behavioural and cognitive complexity [152]; the DOSASE MDP is type 5 of six different types of architecture and is the first type in the spectrum that provides for a developmental capacity. Type 6 builds on this to provide additional attributes, specifically greater abstraction, self-generated contexts, and a higher degree of sensory integration.

The example DOSASE MDP vision system detailed in [151] further elaborates on the cognitive architecture, detailing three types of mapping in the information flow within the architecture: sensory mapping, cognitive mapping, and motor mapping. It is significant that there is more than one cognitive pathway between the sensory mapping and the motor mapping, one of which encapsulates innate behaviours (and the phylogically-endowed capabilities of the system) while the other encapsulates learned behaviours (and the ontogenetically-developed capabilities of the system). These two pathways are mediated by a subsumption-based motor mapping which accords higher priority to the ontogenetically-developed pathway. A second significant feature of the architecture is that it facilitates what Weng refers to as "primed sensations" and "primed action". These correspond to predicitive sensations and actions and thereby provide the system with the anticipative and prospective capabilities that are the hallmark of cognition.

The general SASE schema, including the associated concept of Autonomous Mental Development (AMD), has been developed and validated in the context of two autonomous developmental robotics systems, SAIL and DAV [151], [152], [154], [155].

J. Darwin: Neuromimetic Robotic Brain-Based Devices

Kirchmar *et al.* [16], [156]–[160] have developed a series of robot platforms called Darwin to experiment with developmental agents. These systems are 'brain-based devices' BBDs which that exploit a simulated nervous system that can develop spatial and episodic memory as well as recognition capabilities through autonomous experiential learning. As such, BDDs are a neuromimetic approach in the emergent paradigm that is most closely aligned with the enactive and the connectionist models. It differs from most connectist approaches in that the architecture is much more strongly modelled on the structure and organization of the brain than are conventional artificial neural networks, *i.e.* they focus on the nervous system as a whole, its constituent parts, and their interaction, rather than on a neural implementation of some individual memory, control, or recognition function.

The principal neural mechanisms of the BDD approach are synaptic plasticity, a reward (or value) system, reentrant connectivity, dynamic synchronization of neuronal activity, and neuronal units with spatiotemporal response properties. Adaptive behaviour is achieved by the interaction of these neural mechanisms with sensorimotor correlations (or contingencies) which have been learned autonomously by active sensing and self-motion.

Darwin VIII is capable of discriminating reasonably simple visual targets (coloured geometric shapes) by associating it with an innately preferred auditory cue. Its simulated nervous system contains 28 neural areas, approximately 54,000 neuronal units, and approximately 1.7 million synaptic connections. The architecture comprises regions for vision (V1, V2, V4, IT), tracking (C), value or saliency (S), and audition (A). Gabor filtered images, with vertical, horizontal, and diagonal selectivity, and red-green colour filters with on-centre off-surround and off-centre on-surround receptive fields, are fed to V1. Sub-regions of V1 project topographically to V2 which in turn projects to V4. Both V2 and V4 have excitatory and inhibitory reentrant connections. V4 also has a non-topographical projection back to V2 as well as a non-topographical projection to IT, which itself has reentrant adaptive connections. IT also projects non-toographically back to V4. The tracking area (C) determines the gaze direction of Darwin VIII's camera based on excitatory projections from the auditory region A. This causes Darwin to orient toward a sound source. V4 also projects topographically to C causing Darwin VIII to centre its gaze on a visual object. Both IT and the value system S have adaptive connections to C which facilitates the learned target selection. Adaptation is effected using the Hebbian-like Bienenstock-Cooper-Munroe (BCM) rule [161]. From a behavioural perspective, Darwin VIII is conditioned to prefer one target over others by associating it

with the innately peferred auditory cue and to demonstrate this preference by orienting towards the target.

Darwin IX can navigate and categorize textures using artificial whiskers based on a simulated neuroanatomy of the rat somatosensory system, comprising 17 areas, 1101 neuronal units, and approximately 8400 synaptic connections.

Darwin X is capable of developing spatial and episodic memory based on a model of the hippocampus and surrounding regions. Its simulated nervous system contains 50 neural areas, 90,000 neural units, and 1.4 million synaptic connections. It includes a visual system, head direction system, hippocampal formation, basal forebrain, a value/reward system based on dopaminegic function, and an action selection system. Vision is used to recognize objects and then compute their position, while odometry is used to develop head direction sensitivity.

K. A Humanoid Robot Cognitive Architecture

Burghart *et al.* [162] present a hybrid cognitive architecture for a humanoid robot. It is based on interacting parallel behaviour-based components, comprising a three-level hierarchical perception sub-system, a three-level hierarchical task handling system, a long-term memory sub-system based on a global knowledge database (utilizing a variety of representational schemas, including object ontologies and geometric models, Hidden Markov Models, and kinematic models), a dialogue manager which mediates between perception and task planning, an execution supervisor, and an 'active models' short-term memory sub-system to which all levels of perception and task management have access. These active models play a central role in the cognitive architecture: they are initialized by the global knowledge database and updated by the perceptual sub-system and can be autonomously actualized and reorganized. The perception sub-system comprises a three-level hierarchy with low, mid, and high level perception modules. The low-level perception module provides sensor data interpretation without accessing the central system knowledge database, typically to provide reflex-like low-level robot control. It communicates with both the mid-level perception module and the task execution module. The mid-level perception module provides a variety of recognition components and communicates with both the system knowledge database (long-term memory) as well as the active models (short-term memory). The high-level perception module provides more sophisticated interpretation facilities such as situation recognition, gesture interpretation, movement interpretation, and intention prediction.

The task handling sub-system comprises a three-level hierarchy with task planning, task coordination, and task execution levels. Robot tasks are planned on the top symbolic level using task knowledge. A symbolic plan consists of a set of actions, represented either by XML-files or Petri nets, and acquired either by learning (*e.g.* through demonstration) or by programming. The task planner interacts with the high-level perception module, the (long-term memory) system knowledge database, the task coordination level, and an execution supervisor. This execution supervisor is responsible for the final scheduling of the tasks and resource management in the robot using Petri nets. A sequence of actions is generated and passed down to the task coordination level which then coordinates (deadlock-free) tasks to be run a the lowest task execution (control) level. In general, during the execution of any given task, the task coordination level works independently of the task planning level.

A dialogue manager, which coordinates communication with users and interpretation of communication events, provides a bridge between the perception sub-system and the task sub-system. Its operation is effectively cognitive in the sense that it provides the functionality to recognize the intentions and behaviours of users.

A learning sub-system is also incorporated with the robot currently learning tasks and action sequences off-line by programming by demonstration or tele-operation; on-line learning based on imitation are envisaged. As such, this key component represents work in progress.

L. The Cerebus Architecture

Horswill [163], [164] argues that classical artificial intelligence systems such as those in the tradition of Soar, ART-R, and EPIC, are not well suited for use with robots. Traditional systems typically store all knowledge centrally in a symbolic database of logical assertions and reasoning is concerned mainly with searching and sequentially updating that database. However, robots are distributed systems with multiple sensory, reasoning, and motor control proceses all running in parallel and often only loosely coupled with one another. Each of these processes maintains its own separate and limited representation of the world and the task at hand and he argues that it is not realistic to require them to constantly synchronize with a central knowledge base.

Recently, much the same argument has been made by neuroscientists about the structure and operation of the brain. For example, evidence suggest that space perception is not the result of a single circuit, and in fact derives from the joint activity of several fronto-parietal circuits, each of which encodes the spatial location and transforms it into a potential action in a distinct and motor-specific manner [137], [138]. In other words, the brain encodes space not in a single unified manner — there is no general purpose space map — but in many different ways, each of which is specifically concerned with a particular motor goal. Different motor effectors need different sensory input: derived in different ways and differently encoded in ways that are particular to the different effectors. Conscious space perception emerges from these different pre-existing spatial maps.

Horswill contends also that the classical reasoning systems don't have any good way of directing perceptual attention: they either assume that all the relevant information is already stored in the database or they provide a set of actions that fire task-specific perceptual operators to update specific parts of the database (just as, for example, happens in ACT-R). Both of these approaches are problematic: the former fall foul of the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new

data) and the second requires that the programmer design the rule based to ensure that the appropriate actions are fired in the right circumstances and at the right time; see also similar arguments by Christensen and Hooker [114].

Horswill argues that keeping all of the distinct models or representations in the distributed processes or sub-systems consistent needs to be a key focus of the overall architecture and that is should be done without sychronizing with a central knowledge base. They propose a hybrid cognitive architecture, Cerebus, that combines the tenets of behaviour-based architectures with some features of symbolic AI (forward- and backward-chaining inference using predicate logic). It represents an attempt to scale behaviour-based robots (*e.g.* see Brooks [143] and Arkin [165]) without resorting to a traditional central planning system. It combines a set of behaviour-based sensory-motor systems with a marker-passing semantic network and an inference network. The semantic network effects long-term declarative memory, providing reflective knowledge about its own capabilities, and the inference network allows it to reason about its current state and control processes. Together they implement the key feature of the Cerebus architecture: the use of reflective knowledge about its perceptual-motor systems to perform limited reasoning about its own capabilities.

M. Cog: Theory of Mind

Cog [166] is an upper-torso humanoid robot platform for research on developmental robotics. Cog has a pair of six degree-of-freedom arms, a three degree-of-freedom torso, and a seven degree-of-freedom head and neck. It has a narrow and wide angle binocular vision system (comprising four colour cameras), an auditory system with two microphones, a three-degree of freedom vestibular system, and a range of haptic sensors.

As part of this project, Scassellati has put forward a proposal for a Theory of Mind for Cog [167] that focusses on social interaction as a key aspect of cognitive function in that social skills require the attribution of beliefs, goals, and desires to other people.

A robot that possesses a theory of mind would be capable of learning from an observer using normal social signals and would be capable of expressing its internal state (emotions, desires, goals) though social (non-linguistic) interactions. It would also be capable of recognizing the goals and desires of others and, hence, would be able to anticipate the reactions of the observer and modify its own behaviour accordingly.

Scassellati's proposed architecture is based on Leslie's model of Theory of Mind [168] and Baron-Cohen's model of Theory of Mind [169] both of which decompose the problem into sets of precursor skills and developmental modules, albeit in a different manner. Leslie's Theory of Mind emphasizes independent domain specific modules to distinguish (a) mechanical agency, (b) actional agency, and (c) attitudinal agency; roughly speaking the behaviour of inanimate objects, the behaviour of animate objects, and the beliefs and intentions of animate objects. Baron-Cohen's Theory of Mind comprises three four modules, one of which is concerned with the interpretation of perceptual stimuli (visual, auditory, and tactile) associated with self-propelled motion, and one of which is concerned with the interpretation of visual stimuli associated with eye-like shapes. Both of these feed a shared attention module which in turn feed a Theory of Mind module that represents intentional knowledge or 'epistemic mental states' of other agents.

The focus Scassellati's Theory of Mind for Cog, at least initially, is on the creation of the precursor perceptual and motor skills upon which more complex theory of mind capabilities can be built: distinguishing between inanimate and animate motion and identifying gaze direction. These exploit several built-in visual capabilities such as colour saliency detection, motion detection, skin colour detection, and disparity estimation, a visual search and attention module, and visuo-motor control for saccades, smooth-pursuit, vestibular-ocular reflex, as well as head and neck movement and reaching. The primitive visuo-motor behaviours, *e.g.* for finding faces and eyes, are based on embedded motivational drives and visual search strategies.

N. Kismet

The role of emotion and expressive behaviour in regulating social interaction between humans and robots has been examined by Breazeal using an articulated anthropomorphic robotic head called Kismet [170], [171]. Kismet has a total of 21 degree-of-freedom, three to control the head orientation, three to direct the gaze, and fifteen to control the robots facial features (*e.g.* eye-lids, eyebrows, lips, and ears). Kismet has a narrow and wide angle binocular vision system (comprising four colour cameras), and two microphones, one mounted in each ear. Kismet is designed to engage people in natural and expressive face-to-face interaction, perceiving a natural social cues and responding through gaze direction, facial expression, body posture, and vocal babbling.

Breazeal argues that emotions provide an important mechanism for modulating system behaviour in response to environmental and internal states. They prepare and motivate a system to respond in adaptive ways and serve as reinforcers in learning new behaviour, and act as a mechanism for behavioural homeostasis. The ultimate goal of Kismet is to learn from people though social engagement, although Kismet does not yet have any adaptive (*i.e.* learning or developmental) or anticipatory capabilites.

Kismet has two types of motivations: drives and emotions. Drives establish the top-level goals of the robot: to engage people (social drive), to engage toys (stimulation drive), and to occasionally rest (fatigue drive). The robot's behaviour is focussed on satiating its drives. These drives have a longer time constant compared with emotions. and they operate cyclically: increasing in the absence of satisfying interaction and diminishing with habituation. The goal is to keep the drive level somewhere in a homeostatic region between under stimulation and over stimulation. Emotions — anger & frustration, disgust, fear & distress, calm, joy, sorrow, surprise, interest, boredom — elicit specific behavioural responses such as complain, withdraw, escape, display pleasure, display sorrow, display startled response, re-orient, and seek, in effect tending

to cause the robot to come into contact with things that promote its "well-being" and avoid those that don't. Emotions are triggered by pre-specified antecedent conditions which are based on perceptual stimuli as well as the current drive state and behavioural state.

Kismet has five distinct modules in its cognitive architecture: a perceptual system, an emotion system, a behaviour system, a drive system, and a motor system (see Figure 6).

The perceptual system comprises a set of low-level processes which sense visual and auditory stimuli, perform feature extraction (*e.g.* colour, motion, frequency), extract affective descriptions from speech, orient visual attention, and localize relevant features such as faces, eyes, objects, *etc.*. These are input to a high level perceptual system where, together with affective input from the emotion system, input from the drive system and the behaviour system, they are bound by *releaser* processes 'that encode the robot's current set of beliefs about the state of the robot and its relation to the world. There are many different kinds of releasers, each of which is 'hand-crafted' by the system designer. When the activation level of a releaser exceeds a given threshold (based on the perceptual, affective, drive, and behavioural inputs) it is output to the emotion system for appraisal. Breazeal says that 'each releaser can be thought of as a simple "cognitive" assessment that combines lower-level perceptual features with measures of its internal state into behaviorally significant perceptual categories' [171]. The appraisal process tags the releaser output with pre-specified (*i.e.* designed-in) affective information on their arousal (how much it stimulates the system), valence (how much it is favoured), and stance (how approachable it is). These are then filtered by 'emotion elicitor' to map each AVS (arousal, valence, stance) triple onto the individual emotions. A single emotion is then selected by a winner-take-all arbitration process, and output to the behaviour system and the motor system to evoke the appropriate expression and posture.

Kismet is a hybrid system in the sense that it uses quintessentially cognitivist rule-based schemas to determine, *e.g.*, the antecedent conditions, the operation of the emotion releasers, the affective appraisal, *etc.* but allows the system behaviour to emerge from the dynamic interaction between these sub-systems.

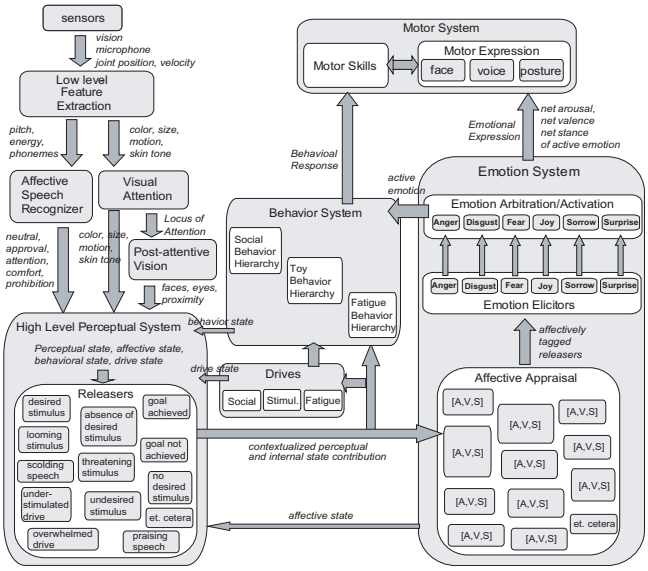

Fig. 6. The Kismet cognitive architecture (from [171]).

Architecture	Paradigm	Embodiment	Perception	Action	Anticipation	Adaptation	Motivation	Autonomy
Soar	C				+	+		
Epic	C		+	+	+			
ACT-R	C		+	+	+	+		
ICARUS	C		+	+	+	+		
ADAPT	C	×	×	×	+	+		
AAR	E	×	×	×			+	×
Global Workspace	E	+	+	+	×		×	×
I-C SDAL	E	+	+	+	+	+	×	×
SASE	E	×	×	×	+	×	×	×
Darwin	E	×	×	+		×	×	×
HUMANOID	H	×	×	×	×	+	+	
Cerebus	H	×	×	×	+	+		
Cog: Theory of Mind	H	×	×	×	+			
Kismet	H	×	×	×			×	

TABLE III
COGNITIVE ARCHITECTURES *vis-à-vis* THE SEVEN OF THE TWELVE CHARACTERISTICS OF COGNITIVE SYSTEMS.

IV. COMPARISON

Table III shows a summary of all the architectures reviewed *vis-à-vis* a subset of the twelve characteristics of cognitive systems which we discussed in Section II. We have omitted the first five characteristics — Computation Operation, Representational Framework, Semantic Grounding, Temporal Constraints, and Inter-agent Epistemology — because these can be inferred directly by the paradigm in which the system is based: cognitivist, emergent, or hybrid, denoted by a C, E, or H in in Table III. A '×' indicates that the characteristic is strongly addressed in the architecture, '+' indicates that it is weakly addressed, and a space indicates that it is not addressed at all in any substantial manner. A '×' is assigned under the heading of Adaptation only if the system is capable of development (in the sense of creating new representational frameworks or models) rather than simple learning (in the sense of model parameter estimation) [151].

V. THE DEVELOPMENTAL STANCE: AUTONOMY, ADAPTATION, LEARNING, AND MOTIVATION

1) Development: Development implies the progressive acquisition of predictive anticipatory capabilities by a system over its lifetime through experiential learning. As we have seen, development requires some ground from which to develop — a phylogenetic configuration — as well as motivations to drive the development.

In the emergent paradigm, the phylogeny must facilitate the autonomy of the system and, in particular, the coupling of the system with its environment, through perception and action, and the self-organization of the system as a distinct en-

tity. This complementary perception/action coupling and self-organization is termed co-determination. Co-determination arises from the autonomous nature of a cognitive system and it reflects the fact that an autonomous system defines itself through a process of self-organization and subjugates all other processes to the preservation of that autonomy [101]. However, it also reflects the fact that all self-organizing systems have an environment in which they are embedded, from which they make themselves distinct, and which is conceived by the autonomous system in whatever way is supportive of this autonomy-preserving process. In this way, the system and the environment are co-specified: the cognitive agent is determined by its environment by its need to sustain its autonomy in the face of environmental perturbations and at the same time the cognitive process determines what is real or meaningful for the agent, for exactly the same reason. In a sense, co-determination means that the agent constructs its reality (its world) as a result of its operation in that world.

Maturana and Varela introduced a diagrammatic way of conveying the self-organized autonomous nature of a co-determined system, perturbing and being perturbed by its environment [45]: see figure 7. The arrow circle denotes the autonomy and self-organization of the system, the rippled line the environment, and the bi-directional half-arrows the mutual perturbation.

Fig. 7. Maturana and Varela's ideograms to denote autopoietic and operationally-closed systems. These systems exhibit co-determination and self-development, respectively. The diagram on the left denotes an autopoietic system: the arrow circle denotes the autonomy, self-organization, and self-production of the system, the rippled line the environment, and the bi-directional half-lines the mutual perturbation —structural coupling —between the two. The diagam on the right denotes an operationally-closed autonomous system with a central nervous system. This system is capable of development by means of self-perturbation —self-modification —of its the nervous system, so that it can accommodate a much larger space of effective system action.

Co-determination requires then that the system is capable of being autonomous as an entity. That is, it has a self-organizing process that is capable of coherent action and perception: that it possesses the essentials of survival and development. This is exactly what we mean by the phylogenetic configuration of a system: the innate capabilities of an autonomous system with which it is equipped at the outset. This, then, forms the ground for subsequent self-development. A co-determined autonomous system has a restricted range of behavioural capabilities and hence a limited degree of autonomy.

Self-development is identically the cognitive process of establishing and enlarging the possible space of mutually-consistent couplings in which a system can engage or withstand whilst maintaining (or increasing) its autonomy. It is the development of the system over time in an ecological and social context as it expands its space of structural couplings that nonetheless must be consistent with the maintenance of self-organization. Self-development requires additional plasticity of the self-organizational processes. The space of perceptual possibilities is predicated not on an absolute objective environment, but on the space of possible actions that the system can engage in whilst still maintaining the consistency of the coupling with the environment. These environmental perturbations don't control the system since they are not components of the system (and, by definition, don't play a part in the self-organization) but they do play a part in the ontogenetic development of the system. Through this ontogenetic development, the cognitive system develops its own epistemology, *i.e.* its own system-specific history- and context-dependent knowledge of its world, knowledge that has meaning exactly because it captures the consistency and invariance that emerges from the dynamic self-organization in the face of environmental coupling. Put simply, the system's actions define its perceptions but subject to the strong constraints of continued dynamic self-organization. Again, it comes down to the preservation of autonomy, but this time doing so in an every increasing space of autonomy-preserving couplings.

This process of development is achieved through self-modification by virtue of the presence of a central nervous system: not only does environment perturb the system (and *vice versa*) but the system also perturbs itself and the central nervous system adapts as a result. Consequently, the system can develop to accommodate a much larger space of effective system action. This is captured in a second ideogram of Maturana and Varela (see figure 7) which adds a second arrow circle to the autopoiesis ideogram to depict the process of self-perturbation and self-modification.

Self-development and co-determination together correspond to Thelen's view that perception, action, and cognition form a single process of self-organization *in the specific context of environmental perturbations of the system* [172]. Thus, we can see that, from this perspective, cognition is inseparable from 'bodily action' [172]: without physical embodied exploration, a cognitive system has no basis for development. Emergent systems, by definition, must be embodied and embedded in their environment in a situated historical developmental context [12].

It is important to emphasize that development occurs in a very special way. Action, perception, and cognition are tightly coupled in development: not only does action organize perception and cognition, but perception and cognition are also essential for organizing action. Actions systems do not appear ready-made. Neither are they primarily determined by experience. They result from both the operation of the central nervous system and the subject's dynamic interactions with the environment. Perception, cognition, and motivations develop at the interface between brain processes and actions. Consequently, cognition can be viewed as the result of a developmental process through which the system becomes

progressively more skilled and acquires the ability to understand events, contexts, and actions, initially dealing with immediate situations and increasingly acquiring a predictive or prospective capability. This dependency on exploration and development is one of the reasons why some argue that the embodied system requires a rich space of manipulation and locomotion actions [47].

We note in passing that the concept of co-determination is rooted in the Maturana's and Varela's idea of structural coupling of level one autopoietic systems[14] [45], is similar to Kelso's circular causality of action and perception each a function of the other as the system manages its mutual interaction with the world [13], and reflect's the organizational principles inherent in Bickhard's self-maintenant systems [14]. The concept of self-development is mirrored in Bickhard's concept of recursive self-maintenance [14] and has its roots in Maturana's and Varela's level two and level three autopoietic systems [45].

In summary, the development of action and perception, the development of the nervous system, and the development (growth) of the body, all mutually influence each other as increasingly-sophisticated and increasingly prospective (future-oriented) capabilities in solving action problems are learned [173].

2) Learning and Motivation: Development depends crucially on motivations which define the goals of actions. The two most important motives that drive actions and development are social and explorative. Social motives include comfort, security, and satisfaction. There are at least two exploratory motives, one involving the discovery of novelty and regularities in the world, and one involving finding out about the potential of one's own actions.

Expanding one's repertoire of actions is a powerful motivation, overriding efficacy in achieving a goal (*e.g.* the development of bi-pedal walking, and the retention of head motion in gaze even in circumstances when ocular control would be more effective). Equally, the discovery of what objects and events afford in the context of new actions is a strong motivation.

The view that exploration is crucial to ontogenetic development is supported by research findings in developmental psychology. For example, von Hofsten has pointed out that it isn't necessarily success at achieving task-specific goals that drives development in neonates but rather the discovery of new modes of interaction: the acquisition of a new way of doing something through exploration [173], [174]. In order to facilitate exploration of new ways of doing things, one must suspend current skills. Consequently, ontogenetic development differs from learning in that (a) it must inhibit existing abilities, and (b) it must be able to cater for (and perhaps effect) changes in the morphology or structure of the system [175]. The inhibition does not imply a loss of learned control but an inhibition of the link between a specific sensory stimulus and a corresponding motor response.

[14]Autopoiesis is a special type of self-organization: an autopoietic system is a homeostatic system (*i.e.* self-regulating system) but one in which the regulation applies not to some system parameter but to the organization of the system itself [45], [101].

In addition to the development of skills through exploration (reaching, grasping, and manipulating what's around it), there are two other very important ways in which cognition develops. These are imitation [176], [177] and social interaction, including teaching [178].

Unlike other learning methods such as reinforcement learning, imitation — the ability to learn new behaviours by observing the actions of others — allows rapid learning [177]. Metzoff and Moore [179], [180] suggest that infants learn through imitation in four phases:

1) body babbling, involving playful trial-and-error movements;
2) imitation of body movements;
3) imitation of actions on objects;
4) imitation based on inferring intentions of others.

Neonates use body babbling to learn a rich "act space" in which new body configurations can be interpolated although its significant that even at birth newborn infants can imitiate body movements [177]. The developmental progress of imitation follows tightly that of the development of other interactive and communicative skills, such as joint attention, turn taking and language [181]–[183]. Imitation is one of the key stages in the development of more advanced cognitive capabilities.

It is important to understand what exactly we mean here by the term 'interaction'. Interaction is a shared activity in which the actions of each agent influence the actions of the other agents engaged in the same interaction, resulting in a mutually constructed pattern of shared behavior [184]. This definition is consistent with the emergent cognition paradigm discussed above, especially the co-constructed nature of the interaction, inspired by concepts of autopoiesis and structural coupling [100]. This aspect of mutually constructed patterns of complementary behavior is also emphasized in Clark's notion of joint action [185]. According to this definition explicit meaning is not necessary for anything to be communicated in an interaction, it is simply important that the agents are mutually engaged in a sequence of actions. Meaning emerges through shared consensual experience mediated by interation.

Development and motivation aside, mechanisms to effect self-modification — or learning — are still required.

Three types of learning can be distinguished: supervised learning in which the teaching signals are directional error signals, reinforcement learning in which the teaching signals are scalar rewards or reinforcement signals, and unsupervised learning with no teaching signals. Doya argues that the cerebellum is specialized for supervised learning, basal ganglia for reinforcement learning, and the cerebral cortex for unsupervised learning [186]. He suggests that in developing (cognitive) architectures, the supervised learning modules in the cerebellum can be used as an internal model of the environment and as short-cut models of input-output mappings that have been acquired elsewhere in the brain. Reinforcement learning modules in the basal ganglia are used to evaluate a given state and thereby to select an action. The unsupervised modules in the cerebral cortex represent the state of the external environment as well as internal context, providing also a common representational framework for the cerebellum and the basal ganglia which have no direct anatomical connections.

Irrespective of the exact details of Doya's model, what is significant is that different regions facilitate different types of learning and that these regions and the learning processes are interdependent. For example, McClelland *et al.* have suggested that the hippocampal formation and the neo-cortex form a complementary system for learning [187]. The hippocampus facilitates rapid auto- and hetero-associative learning which is used to reinstate and consolidate learned memories in the neo-cortex in a gradual manner. In this way, the hippocampal memory can be viewed not just as a memory store but as a 'teacher of the neo-cortical processing system'. Note also that the reinstatement can occur on-line, thereby enabling the overt control of behavioural responses, as well as off-line in, *e.g.* active rehearsal, reminiscence, and sleep.

In a similar vein, Rougier has proposed and validated an architecture for an auto-associative memory based on the organization of the hippocampus, involving the entorhinal cortex, the dentate gyrus, CA3, and CA1 [188]. A feature of this architecture is that it avoids the catastrophic interference problem normally linked to associative memories through the use of redundancy, orthogonalization, and coarse coding representations. Rougier too notes that the hippocampus plays a role in 'teaching' the neo-cortex, *i.e.* in the formation of neocortical representations.

Different types of development require different learning mechanisms. Innate behaviours are honed through continuous knowledge-free reinforcement-like learning in a process somewhat akin to parameter estimation. On the other hand, new skills develop through a different form of learning, driven not just by conventional reward/punishment cost functions (positive and negative feedback) but through spontaneous unsupervised play and exploration which are not directly reinforced [189], [190].

In summary, cognitive skills emerge progressively through ontogenetic development as it learns to make sense of its world through exploration, through manipulation, imitation, and social interaction, including communication [47]. Proponents of the enactive approach would add the additional requirement that this development take place in the context of a circular causality of action and perception, each a function of the other as the system manages its mutual interaction with the world: essentially self-development of action and perception, and co-determination of the system through self-organization in an ecological and social context.

To conclude, Winograd and Flores [24] capture the essence of developmental emergent learning very succinctly:

> 'Learning is not a process of accumulation of representations of the environment; it is a continuous process of transformation of behaviour through continuous change in the capacity of the nervous system to synthesize it. Recall does not depend on the indefinite retention of a structural invariant that represents an entity (an idea, image, or symbol), but on the functional ability of the system to create, when certain recurrent conditions are given, a behaviour that satisfies the recurrent demands or that the observer would class as a reenacting of a previous one'.

3) Perception/Action Co-Dependency: An Example of Self-Development: It has been shown that perception and action in biological systems are co-dependent. For example, spatial attention is dependent on oculomotor programming: when the eye is positioned close to the limit of its rotation, and therefore cannot saccade in any further in one direction, visual attention in that direction is attenuated [191]. This premotor theory of attention applies not only to spatial attention but also to selective attention in which some object rather than others are more apparent. For example, the ability to detect an object is enhanced when features or the appearance of the object coincide with the grasp configuration of a subject preparing to grasp an object [192]. In other words, the subject's actions conditions its perceptions. Similarly, the presence of a set of neurons — mirror neurons — is often cited as evidence of the tight relationship between perception and action [193], [194]. Mirror neurons are activated both when an action is performed and when the same or similar action is observed being performed by another agent. These neurons are specific to the goal of the action and not the mechanics of carrying it out [173]. Furthermore, perceptual development is determined by the action capabilities of a developing child and on what observed objects and events afford in the context of those actions [173], [195].

A practical example of a system which exploits this co-dependency in a developmental setting can be found in [87]. This is a biologically-motivated system that learns goal-directed reaching using colour-segmented images derived from a retina-like log-polar sensor camera. The system adopts a developmental approach: beginning with innate inbuilt primitive reflexes, it learns sensorimotor coordination. The system operates as follows. By assuming that a fixation point represents the object to be reached for, the reaching is effected by mapping the eye-head proprioceptive data to the arm control parameters. The control itself is implemented as a multi-joint synergy by using the control parameters to modulate a linear combination of basis torque fields, each torque field describing the torque to be applied to an actuator or group of actuators to achieve some distinct equilibrium point where the acuator position is stable. That is, the eye-hand motor commands which direct the gaze towards a fixation point are used to control the arm motors, effecting what is referred to in the paper as "motor-motor coordination". The mapping between eye-head proprioceptive data (joint angular positions) and the arm control parameters is learned by fixating on the robot hand during a training phase.

A similar but more extensive biologically-motivated system, modelled on brain function and cortical pathways and exploiting optical flow as its primary visual stimulus, demonstrates the development of object segmentation, recognition, and localization capabilities without any prior knowledge of visual appearance though exploratory reaching and simple manipulation [112]. The system also exhibits the ability to learn a simple object affordance and use it to mimic the actions of another (human) agent. The working hypothesis is that action is required for object recognition in cases where the system has to develop the object classes or categories autonomously. The inherent ambiguity in visual perception can be resolved by acting upon the environment that is perceived. Development

starts with reaching, and proceeds through grasping, and ultimately to object recognition. Training the arm-gaze controller is effected in much the same way as in [87] but in this case, rather than using colour segmentation, the arm is segmented by seeking optical flow that is correlated with arm movements (specifically, during training, by correlating discontinuities in arm movement as it changes direction of motion with temporal discontinuities in the flow field. Segmentation of (movable) objects is effected also by optical flow by poking the object and detecting regions in the flow field that are also correlated with arm motion, but which can't be attributed to the arm itself. Objects that are segmented by poking can them be classified using colour histograms of the segmented regions. A simple affordance — rolling behaviour when poked — is learned by computing the probability of a normalized direction of motion when the object is poked (normalization is effected by taking the difference between the principal axis of the object and the angle of motion). The effect of different poking gestures on objects is then learned for each gesture by computing the probability density function (a histogram, in effect) of the direction of motions averaged over all objects. There are four gestures in all: pull in, push away, backslap, and side tap. When operating in a non-exploratory mode, object recognition is effected by colour histogram matching, localization by histogram back-projection, and orientation by estimating the principal axis by comparison of the segmented object with learned prototypes. The robot then selects an action (one of the four gestures) by finding the preferred rolling direction (from its learned affordances) adding it to the current orientation and then choosing the gesture which has the highest probability associated with resultant direction. Mimicry (which differs from imitation, the latter being associated with learning new behaviour, and the former with repeating known behaviour [176]) is effected by presenting the robot with an object and performing an action on it. This "action to be imitated" activity is flagged by detecting motion in the neighbourhood of the fixation point, reaching by the robot is then inhibited, and the effect of the action of the object is observed using optical flow and template matching. When the object is presented again a second time, the poking action that is most likely to reproduce the rolling affordance is selected. It is assumed that this is exactly what one would expect of a mirror-neuron type of representation of perception and action. Mirror neurons can be thought of as an "associative map that links together the observation of a manipulative action performed by someone else with the neural representation of one's own actions".

VI. Implications for the Autonomous Development of Mental Capabilities in Computational Systems

We finish this survey by drawing together the main issues raised in the foregoing and we summarize some of the key features that a system capable of autonomous mental development, *i.e.* an artificial cognitive system, should exhibit, especially those that adhere to a developmental approach. However, before doing this, it might be opportune to remark first on the dichotomy between cognitivist and emergent systems. As we have seen, there are some fundamental differences these two general paradigms — the principalled disembodiment of physical symbol systems *vs.* the mandatory embodiment of emergent developmental systems [48], and the manner in which cognitivist systems often preempt development by embedding externally-derived domain knowledge and processing structures, for example — but the gap between the two shows some signs of narrowing. This is mainly due (i) to a fairly recent movement on the part of proponents of the cognitivist paradigm to assert the fundamentally important role played by action and perception in the realization of a cognitive system; (ii) to the move away from the view that internal symbolic representations are the only valid form of representation [10]; and (iii) to the weakening of the dependence on embedded *a priori* knowledge and the attendant increased reliance on machine learning and statistical frameworks both for tuning system parameters and the acquisition of new knowledge both for the representation of objects and the formation of new representations. However, cognitivist systems still have some way to go to address the issue of true ontogenetic development with all that it entails for autonomy, embodiment, architecture plasticity, and system-centred construction of knowledge mediated by exploratory and social motivations and innate value systems.

Krichmar *et al.* identify six design principles for systems that are capable of development [16], [156], [159]. Although they present these principles in the context of their brain-based devices, most are directly applicable to emergent systems in general. First, they suggest that the architecture should address the dynamics of the neural element in different regions of the brain, the structure of these regions, and especially the connectivity and interaction between these regions. Second, they note that the system should be able to effect perceptual categorization: *i.e.* to organize unlabelled sensory signals of all modalities into categories without *a priori* knowledge or external instruction. In effect, this means that the system should be autonomous and, as noted by Weng [151], p. 206, a developmental system should be a model generator, rather than a model fitter (*e.g.* see [196]). Third, a developmental system should have a physical instantiation, *i.e.* it should be embodied, so that it is tightly coupled with its own morphology and so that it can explore its environment. Fourth, the system should engage in some behavioural task and, consequently, it should have some minimal set of innate behaviours or reflexes in order to explore and survive in its initial environmental niche. From this minimum set, the system can learn and adapt so that it improves[15] its behaviour over time. Fifth, developmental systems should have a means to adapt. This implies the presence of a value system (*i.e.* a set of motivations that guide or govern its development). These should be non-specific[16] modulatory signals that bias the dynamics of the system so that the global needs of the system are satisfied: in effect, so that its autonomy is preserved or enhanced. Such value systems might possibly be modelled on the value system of the brain: dopaminergic, cholinergic, and noradrenergic

[15]Krichmar *et al.* say 'optimizes' rather than 'improves'.

[16]Non-specific in the sense that they don't specify what actions to take.

systems signalling, on the basis of sensory stimuli, reward prediction, uncertainty, and novelty. Krichmar *et al.* also note that brain-based devices should lend themselves to comparison with biological systems.

And so, with both the foregoing survey and these design principles, what conclusions can we draw?

First, a developmental cognitive system will be constituted by a network of competing and cooperating distributed multi-functional sub-systems (or cortical circuits), each with its own limited encoding or representational framework, together achieving the cognitive goal of effective behaviour, effected either by some self-synchronizing mechanism or by some modulation circuit. This network forms the system's phylogenetic configuration and its innate abilities.

Second, a developmental cognitive architecture must be capable of adaptation and self-modification, both in the sense of parameter adjustment of phylogenetic skills through learning and, more importantly, through the modification of the very structure and organization of the system itself so that it is capable of altering its system dynamics based on experience, to expand its repertoire of actions, and thereby adapt to new circumstances. This development should be driven by both explorative and social motives, the first concerned with both the discovery of novel regularities in the world and the potential of the system's own actions, the second with inter-agent interaction, shared activities, and mutually-constructed pattern's of shared behaviour. A variety of learning paradigms will need to be recruited to effect development, including, but not necessarily limited to, unsupervised, reinforcement, and supervised learning.

Third, and because cognitive systems are not only adaptive but also anticipatory and prospective, it is crucial that they have (by virtue of their phylogeny) or develop (by virtue of their ontogeny) some mechanism to rehearse hypothetical scenarios — explicitly like Anderson's ACT-R architecture [7] or implicitly like Shanahan's global workspace dynamical architecture [144] — and a mechanism to then use this to modulate the actual behaviour of the system.

Finally, developmental cognitive systems have to be embodied, at the very least in the sense of stuctural coupling with the environment and probably in some stronger organismoid form [197], [198], if the epistemological understanding of the developed systems is required to be consistent with that of other cognitive agents such as humans [3]. What is clear, however, is that the complexity and sophistication of the cognitive behaviour is dependent on the richness and diversity of the coupling and therefore the potential richness of the system's actions.

Ultimately, for both cognitivist and emergent paradigms, development (*i.e.* ontogeny), is dependent on the system's phylogenetic configuration as well as its history of interactions and activity. Exactly what phylogenetic configuration is required for the autonomous development of mental capabilities — *i.e.* for the construction of artificial cognitive systems with mechanisms for perception, action, adaptation, anticipation, and motivation that enable its ontogenetic development over its life-time — remains an open question. Hopefully, this survey will go some way towards answering it.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the many helpful comments of the two anonymous referees on earlier versions of this paper.

REFERENCES

[1] M. L. Anderson, "Embodied cognition: A field guide," *Artificial Intelligence*, vol. 149, no. 1, pp. 91–130, 2003.
[2] A. Berthoz, *The Brain's Sense of Movement*. Cambridge, MA: Harvard University Press, 2000.
[3] D. Vernon, "The space of cognitive vision," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, ser. LNCS (In Press), H. I. Christensen and H.-H. Nagel, Eds. Heidelberg: Springer-Verlag, 2006, pp. 7–26.
[4] R. J. Brachman, "Systems that know what they're doing," *IEEE Intelligent Systems*, vol. 17, no. 6, pp. 67–71, Dec. 2002.
[5] E. Hollnagel and D. D. Woods, "Cognitive systems engineering: New wind in new bottles," *International Journal of Human-Computer Studies*, vol. 51, pp. 339–356, 1999.
[6] W. J. Freeman and R. N´u˜nez, "Restoring to cognition the forgotten primacy of action, intention and emotion," *Journal of Consciousness Studies*, vol. 6, no. 11-12, pp. ix–xix, 1999.
[7] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, pp. 1036–1060, 2004.
[8] P. Langley, "An adaptive architecture for physical agents," in *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*. Compiegne, France: IEEE Computer Society Press, 2005, pp. 18–25.
[9] F. J. Varela, "Whence perceptual meaning? A cartography of current ideas," in *Understanding Origins – Contemporary Views on the Origin of Life, Mind and Society*, ser. Boston Studies in the Philosophy of Science, F. J. Varela and J.-P. Dupuy, Eds. Kluwer Academic Publishers, 1992, pp. 235–263.
[10] A. Clark, *Mindware – An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press, 2001.
[11] Z. W. Pylyshyn, *Computation and Cognition*, 2nd ed. Bradford Books, MIT Press, 1984.
[12] E. Thelen and L. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action*, ser. MIT Press / Bradford Books Series in Cognitive Psychology. Cambridge, Massachusetts: MIT Press, 1994.
[13] J. A. S. Kelso, *Dynamic Patterns – The Self-Organization of Brain and Behaviour*, 3rd ed. MIT Press, 1995.
[14] M. H. Bickhard, "Autonomy, function, and representation," *Artificial Intelligence, Special Issue on Communication and Cognition*, vol. 17, no. 3-4, pp. 111–131, 2000.
[15] W. D. Christensen and C. A. Hooker, "An interactivist-constructivist approach to intelligence: self-directed anticipative learning," *Philosophical Psychology*, vol. 13, no. 1, pp. 5–45, 2000.
[16] J. L. Krichmar and G. M. Edelman, "Principles underlying the construction of brain-based devices," in *Proceedings of AISB '06 - Adaptation in Artificial and Biological Systems*, ser. Symposium on Grand Challenge 5: Architecture of Brain and Mind, T. Kovacs and J. A. R. Marshall, Eds., vol. 2. Bristol: University of Bristol, 2006, pp. 37–42.
[17] H. Gardner, *Multiple Intelligences: The Theory in Practice*. New York: Basic Books, 1993.
[18] W. S. McCulloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.
[19] D. Marr, "Artificial intelligence – A personal view," *Artificial Intelligence*, vol. 9, pp. 37–48, 1977.
[20] A. Newell and H. A. Simon, "Computer science as empirical inquiry: Symbols and search," *Communications of the Association for Computing Machinery*, vol. 19, pp. 113–126, Mar. 1976, tenth Turing award lecture, ACM, 1975.
[21] J. Haugland, "Semantic engines: An introduction to mind design," in *Mind Design: Philosophy, Psychology, Artificial Intelligence*, J. Haugland, Ed. Cambridge, Massachusetts: Bradford Books, MIT Press, 1982, pp. 1–34.
[22] S. Pinker, "Visual cognition: An introduction," *Cognition*, vol. 18, pp. 1–63, 1984.

[23] J. F. Kihlstrom, "The cognitive unconscious," *Science*, vol. 237, pp. 1445–1452, Sept. 1987.

[24] T. Winograd and F. Flores, *Understanding Computers and Cognition – A New Foundation for Design*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1986.

[25] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[26] J. Pauli and G. Sommer, "Perceptual organization with image formation compatibilities," *Pattern Recognition Letters*, vol. 23, no. 7, pp. 803–817, 2002.

[27] R. N. Shepard and S. Hurwitz, "Upward direction, mental rotation, and discrimination of left and right turns in maps," *Cognition*, vol. 18, pp. 161–193, 1984.

[28] H.-H. Nagel, "Steps toward a cognitive vision system," *AI Magazine*, vol. 25, no. 2, pp. 31–50, Summer 2004.

[29] M. Arens and H.-H. Nagel, "Quantitative movement prediction based on qualitative knowledge about behaviour," *KI-Zeitschrift Knstliche Intelligenz, Special Issue on Cognitive Computer Vision*, pp. 5–11, Apr. 2005.

[30] M. Arens and H. H. Nagel, "Representation of behavioral knowledge for planning and plan recognition in a cognitive vision system," in *Proceedings iof the 25th German Conference on Artifi cial Intelligence (KI-2002)*, M. Jarke, J. Koehler, and G. Lakemeyer, Eds. Aachen, Germany: Springer-Verlag, Sept. 2002, pp. 268–282.

[31] M. Arens, A. Ottlick, and H. H. Nagel, "Natural language texts for a cognitive vision system," in *Proceedings of the 15th European Conference On Artifi cial Intelligence (ECAI-2002)*, F. V. Harmelen, Ed. Amsterdam: IOS Press, 2002, pp. 455–459.

[32] R. Gerber, H. H. Nagel, and H. Schreiber, "Deriving textual descriptions of road traffic queues from video sequences," in *Proceeedings of the 15th European Conference on Artifi cial Intelligence (ECAI-2002)*, V. H. F, Ed. Amsterdam: IOS Press, 2002, pp. 736–740.

[33] R. Gerber and N. H. H., "'occurrence' extraction from image sequences of road traffic," in *Cognitive Vision Workshop*, Zurich, Switzerland, Sept. 2002.

[34] B. Neumann and R. M¨oller, "On scene interpretation with description logics," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, ser. LNCS (In Press), H. I. Christensen and H.-H. Nagel, Eds. Heidelberg: Springer-Verlag, 2005, pp. 235–260.

[35] R. M¨oeller, B. Neumann, and M. Wessel, "Towards computer vision with description logics: Some recent progress," in *Proc. Integration of Speech and Image Understanding*. Corfu, Greece: IEEE Computer Society, 1999, pp. 101–115.

[36] H. Buxton, "Generative Models for Learning and Understanding Dynamic Scene Activity," in *ECCV Workshop on Generative Model Based Vision*, Copenhagen, Denmark, June 2002.

[37] K. Sage, J. Howell, and H. Buxton, "Recognition of action, activity and behaviour in the actIPret project," *KI-Zeitschrift Knstliche Intelligenz, Special Issue on Cognitive Computer Vision*, pp. 30–34, Apr. 2005.

[38] H. Buxton and A. J. Howell, "Active Vision Techniques for Visually Mediated Interaction," in *International Conference on Pattern recognition*, Quebec City, Canada, Aug. 2002.

[39] H. Buxton, A. J. Howell, and K. Sage, "The Role of Task Control and Context in Learning to Recognise Gesture," in *Workshop on Cognitive Vision*, Z¨urich, Switzerland, Sept. 2002.

[40] A. G. Cohn, D. C. Hogg, B. Bennett, V. Devin, A. Galata, D. R. Magee, C. Needham, and P. Santos, "Cognitive vision: Integrating symbolic qualitative representations with computer vision," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, ser. LNCS, H. I. Christensen and H.-H. Nagel, Eds. Heidelberg: Springer-Verlag, 2005, pp. 211–234.

[41] N. Maillot, M. Thonnat, and A. Boucher, "Towards ontology based cognitive vision," in *Proceedings of the Third International Conference on Computer Vision Systems, ICVS 2003*, J. Crowley, J. Piater, M. Vincze, and L. Paletta, Eds., vol. LNCS 2626. Berlin Heidelberg: Springer-Verlag, 2003, pp. 44–53.

[42] A. Chella, M. Frixione, and S. Gaglio, "A cognitive architecture for artificial vision," *Artifi cial Intelligence*, vol. 89, no. 1–2, pp. 73–111, 1997.

[43] J. L. Crowley, "Things that see: Context-aware multi-modal interaction," *KI-Zeitschrift Knstliche Intelligenz, Special Issue on Cognitive Computer Vision*, Apr. 2005.

[44] E. D. Dickmanns, "Dynamic vision-based intelligence," *AI Magazine*, vol. 25, no. 2, pp. 10–29, Summer 2004.

[45] H. Maturana and F. Varela, *The Tree of Knowledge – The Biological Roots of Human Understanding*. Boston & London: New Science Library, 1987.

[46] G. H. Granlund, "The complexity of vision," *Signal Processing*, vol. 74, pp. 101–126, 1999.

[47] G. Sandini, G. Metta, and D. Vernon, "Robotcub: An open framework for research in embodied cognition," in *IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids 2004)*, 2004, pp. 13–32.

[48] D. Vernon, "Cognitive vision: The case for embodied perception," *Image and Vision Computing*, vol. In Press, pp. 1–14, 2006.

[49] D. A. Medler, "A brief history of connectionism," *Neural Computing Surveys*, vol. 1, pp. 61–101, 1998.

[50] P. Smolensky, "Computational, dynamical, and statistical perspectives on the processing and learning problems in neural network theory," in *Mathematical perspectives on neural networks*, P. Smolensky, M. C. Mozer, and D. E. Rumelhart, Eds. Erlbaum, 1996, pp. 1–15.

[51] J. A. Anderson and E. Rosenfeld, Eds., *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press, 1988.

[52] ——, *Neurocomputing 2: Directions for Research*. Cambridge, MA: MIT Press, 1991.

[53] P. Smolensky, "Computational perspectives on neural networks," in *Mathematical perspectives on neural networks*, P. Smolensky, M. C. Mozer, and D. E. Rumelhart, Eds. Erlbaum, 1996, pp. 1–15.

[54] ——, "Dynamical perspectives on neural networks," in *Mathematical perspectives on neural networks*, P. Smolensky, M. C. Mozer, and D. E. Rumelhart, Eds. Erlbaum, 1996, pp. 245–270.

[55] ——, "Statistical perspectives on neural networks," in *Mathematical perspectives on neural networks*, P. Smolensky, M. C. Mozer, and D. E. Rumelhart, Eds. Erlbaum, 1996, pp. 453–496.

[56] M. A. Arbib, Ed., *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press, 95.

[57] R. A. Feldman and D. H. Ballard, "Connectionist models and their properties," *Cognitive Science*, vol. 6, pp. 205–254, 1982.

[58] E. L. Thorndike, *The Fundamentals of Learning*. New York: Teachers College, Columbia University, 1932.

[59] ——, *Selected Writings from a Connectionist Psychology*. New York: Greenwood Press, 1949.

[60] W. James, *The Principles of Psychology*, 1890, vol. 1.

[61] D. O. Hebb, *The Organization of Behaviour*. New York: John Wiley & Sons, 1949.

[62] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.

[63] O. G. Selfridge, "Pandemonium: A paradigm for learning," in *Proceedings of the Symposium on Mechanization of Thought Processes*, D. V. Blake and A. M. Uttley, Eds. London: H. M. Stationery Office, 1959, pp. 511–529.

[64] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record*, New York, 1960, pp. 96–104.

[65] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

[66] G. E. Hinton and J. A. Anderson, Eds., *Parallel models of associative memory*. Hillsdale, N.J.:: Lawrence Eralbaum Associates, 1981.

[67] J. L. McClelland, "Retrieving general and specific information from stored knowledge of specifics," in *Proceedings of the Third Annual Meeting of the Cognitive Science Society*, 1981, pp. 170–172.

[68] S. Grossberg, "Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors," *Biological Cybernetics*, vol. 23, pp. 121–134, 1976.

[69] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.

[70] G. A. Carpenter and S. Grossberg, "Adaptive resonance theory (art)," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 1995, pp. 79–82.

[71] D. E. Rumelhart, J. L. McClelland, and The PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge: The MIT Press, 1986.

[72] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, and The PDP Research Group, Eds. Cambridge: The MIT Press, 1986, pp. 318–362.

[73] ——, "Learning representations by back-propagating erros," *Nature*, vol. 323, pp. 533–536, 1986.

[74] P. Werbos, *Beyond regression: new tools for prediction and analysis in the behaviourl sciences*, ser. Masters Thesis. Boston, MA: Harvard University, 131–140 1974.

[75] J. J. Hopfield, "Neural neural network and physical systems with emergent collective computational abilities," *Proceedings of National Academy of Sciences*, vol. 79, no. 8, pp. 2554 – 2588, 1982.

[76] J. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179–211, 1990.

[77] M. I. Jordan, "Attractor dynamics and parallelism in a connectionist sequential machine," in *Proceedings of the Eighth Conference of the Cognitive Science Society*, 1986, pp. 531–546.

[78] G. E. Hinton and T. J. Sejnowski, "Learning and relearning in boltzmann machines," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart, J. L. McClelland, and The PDP Research Group, Eds. Cambridge: The MIT Press, 1986, pp. 282–317.

[79] J. Moody and C. J. Darken, "Fast learning in networks of locally tuned processing units," *Neural Computation*, vol. 1, pp. 281–294, 1989.

[80] J. L. McClelland and T. T. Rogers, "The parallel distributed processing approach to semantic cognition," *Nature*, vol. 4, pp. 310–322, 2003.

[81] P. Smolensky and G. Legendre, *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar*. MIT Press, 2006.

[82] P. Smolensky, "structure and explanation in an integrated connectionist/symbolic cognitive architecture," in *Connectionism: Debates on psychological explanation*, C. Macdonald and G. Macdonald, Eds. Basil Blackwell, 1995, vol. 2, pp. 221–290.

[83] T. van Gelder and R. F. Port, "It's about time: An overview of the dynamical approach to cognition," in *Mind as Motion – Explorations in the Dynamics of Cognition*, R. F. Port and T. van Gelder, Eds. Cambridge, Massachusetts: Bradford Books, MIT Press, 1995, pp. 1–43.

[84] M. Jones and D. Vernon, "Using neural networks to learn hand-eye co-ordination," *Neural Computing and Applications*, vol. 2, no. 1, pp. 2–12, 1994.

[85] B. W. Mel, "MURPHY: A robot that learns by doing," in *Neural Information Processing Systems*. American Institute of Physics, 1988, pp. 544–553.

[86] R. Linsker, "Self-organization in a perceptual network," *Computer*, pp. 105–117, Mar. 1988.

[87] G. Metta, G. Sandini, and J. Konczak, "A developmental approach to visually-guided reaching in artificial systems," *Neural Networks*, vol. 12, no. 10, pp. 1413–1427, 1999.

[88] R. Reiter, *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, Massachusetts: MIT Press, 2001.

[89] J. J. Gibson, *The Perception of the Visual World*. Boston: Houghton Mifflin, 1950.

[90] ——, *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.

[91] W. Köhler, *Dynamics in Psychology*. New York: Liveright, 1940.

[92] W. H. Warren, "Perceiving affordances: Visual guidance of stairclimbing," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 10, pp. 683–703, 1984.

[93] J. L. McClelland and G. Vallabha, "Connectionist models of development: Mechanistic dynamical models with emergent dynamical properties," in *Toward a New Grand Theory of Development? Connectionism and Dynamic Systems Theory Re-Considered*, J. P. Spencer, M. S. C. Thomas, and J. L. McClelland, Eds. New York: Oxford University Press, 2006.

[94] H. R. Wilson, *Spikes, Decisions, and Actions: Dynamical Foundations of Neurosciences*. Oxford University Press, 1999.

[95] G. Schöner, "Development as change of dynamic systems: Stability, instability, and emergence," in *Toward a New Grand Theory of Development? Connectionism and Dynamic Systems Theory Re-Considered*, J. P. Spencer, M. S. C. Thomas, and J. L. McClelland, Eds. New York: Oxford University Press, 2006.

[96] G. Schöner and J. A. S. Kelso, "Dynamic pattern generation in behavioural and neural systems," *Science*, vol. 239, pp. 1513–1520, 1988.

[97] D. Marr, *Vision*. San Francisco: Freeman, 1982.

[98] H. Maturana, "Biology of cognition," University of Illinois, Urbana, Illinois, Research Report BCL 9.0, 1970.

[99] ——, "The organization of the living: a theory of the living organization," *Int. Journal of Man-Machine Studies*, vol. 7, no. 3, pp. 313–332, 1975.

[100] H. R. Maturana and F. J. Varela, *Autopoiesis and Cognition — The Realization of the Living*, ser. Boston Studies on the Philosophy of Science. Dordrecht, Holland: D. Reidel Publishing Company, 1980.

[101] F. Varela, *Principles of Biological Autonomy*. New York: Elsevier North Holland, 1979.

[102] D. Philipona, J. K. O'Regan, and J.-P. Nadal, "Is there something out there? Inferring space from sensorimotor dependencies," *Neural Computation*, vol. 15, no. 9, 2003.

[103] D. Philipona, J. K. O'Regan, J.-P. Nadal, and O. Coenen, "Perception of the structure of the physical world using unknown multimodal sensors and effectors," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2004.

[104] G. H. Granlund, "Does vision inevitably have to be active?" in *Proceedings of SCIA99, Scandanavian Conference on Image Analysis*, 1999.

[105] ——, "Cognitive vision – background and research issues," Linköping University," Research Report, 2002.

[106] H. L. Dreyfus, "From micro-worlds to knowledge representation," in *Mind Design: Philosophy, Psychology, Artificial Intelligence*, J. Haugland, Ed. Cambridge, Massachusetts: Bradford Books, MIT Press, 1982, pp. 161–204, excerpted from the Introduction to the second edition of the author's *What Computers Can't Do*, Harper and Row, 1979.

[107] D. H. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, pp. 57–86, 1991.

[108] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proceeding of the 8th European Conference on Computer Vision, ECCV 2004*, ser. LNCS, T. Pajdla and J. Matas, Eds., vol. 3021. Springer-Verlag, 2004, pp. 28–39.

[109] G. Granlund, "A cognitive vision architecture integrating neural networks with symbolic processing," *KI-Zeitschrift Knstliche Intelligenz, Special Issue on Cognitive Computer Vision*, Apr. 2005.

[110] ——, "Organization of architectures for cognitive vision systems," in *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, ser. LNCS, H. I. Christensen and H.-H. Nagel, Eds. Heidelberg: Springer-Verlag, 2005, pp. 39–58.

[111] G. Granlund and A. Moe, "Unrestricted recognition of 3D objects for robotics using multilevel triplet invariants," *AI Magazine*, vol. 25, no. 2, pp. 51–67, Summer 2004.

[112] G. Metta and P. Fitzpatrick, "Early integration of vision and manipulation," *Adaptive Behavior*, vol. 11, no. 2, pp. 109–128, 2003.

[113] M. Jogan, M. Artac, D. Skocaj, and A. Leonardis, "A framework for robust and incremental self-localization of a mobile robot," in *Proceedings of the Third International Conference on Computer Vision Systems, ICVS 2003*, J. Crowley, J. Piater, M. Vincze, and L. Paletta, Eds., vol. LNCS 2626. Berlin Heidelberg: Springer-Verlag, 2003, pp. 460–469.

[114] W. D. Christensen and C. A. Hooker, "Representation and the meaning of life," in *Representation in Mind: New Approaches to Mental Representation*, The University of Sydney, June 2000.

[115] J. P. Crutchfield, "Dynamical embodiment of computation in cognitive processes," *Behavioural and Brain Sciences*, vol. 21, no. 5, pp. 635–637, 1998.

[116] M. P. Shanahan and B. Baars, "Applying global workspace theory to the frame problem," *Cognition*, vol. 98, no. 2, pp. 157–176, 2005.

[117] G. Metta, D. Vernon, and G. Sandini, "The robotcub approach to the development of cognition: Implications of emergent systems for a common research agenda in epigenetic robotics," in *Proceedings of the Fifth International Workshop on Epigenetic Robotics (EpiRob2005)*, 2005.

[118] A. Newell, "The knowledge level," *Artificial Intelligence*, vol. 18, no. 1, pp. 87–127, Mar. 1982.

[119] ——, *Unified Theories of Cognition*. Cambridge MA: Harvard University Press, 1990.

[120] P. Rosenbloom, J. Laird, and A. Newell, Eds., *The Soar Papers: Research on Integrated Intelligence*. Cambridge, Massachusetts: MIT Press, 1993.

[121] M. D. Byrne, "Cognitive architecture," in *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 97–117.

[122] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: an architecture for general intelligence," *Artificial Intelligence*, vol. 33, no. 1–64, 1987.

[123] J. F. Lehman, J. E. Laird, and P. S. Rosenbloom, "A gentle introduction to soar, an architecture for human cognition," in *Invitation to Cognitive Science, Volume 4: Methods, Models, and Conceptual Issues*, S. Sternberg and D. Scarborough, Eds. Cambridge, MA: MIT Press, 1998.

[124] R. L. Lewis, "Cognitive theory, soar," in *International Encyclopedia of the Social and Behavioural Sciences*. Amsterdam: Pergamon (Elsevier Science), 2001.

[125] J. R. Anderson, "Act: A simple theory of complex cognition," *American Psychologist*, vol. 51, pp. 355–365, 1996.

[126] M. Minsky, *Society of Mind*. New York: Simon and Schuster, 1986.

[127] W. D. Gray, R. M. Young, and S. S. Kirschenbaum, "Introduction to this special issue on cognitive architectures and human-computer interaction," *Human-Computer Interaction*, vol. 12, pp. 301–309, 1997.

[128] F. E. Ritter and R. M. Young, "Introduction to this special issue on using cognitive models to improve interface design," *International Journal of Human-Computer Studies*, vol. 55, pp. 1–14, 2001.

[129] *A Survey of Cognitive and Agent Architectures*, `http://ai.eecs.umich.edu/cogarch0/`.

[130] A. Karmiloff-Smith, *Beyond Modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press, 1992.

[131] ——, "Precis of beyond modularity: A developmental perspective on cognitive science," *Behavioral and Brain Sciences*, vol. 17, no. 4, pp. 693–745, 1994.

[132] J. A. Fodor, *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press, 1983.

[133] S. Pinker, *How the Mind Works*. New York: W. W. Norton and Company, 1997.

[134] J. A. Fodor, *The Mind Doesn't Work that Way*. Cambridge, MA: MIT Press, 2000.

[135] J. Piaget, *The Construction of Reality in the Child*. London: Routeledge and Kegan Paul, 1955.

[136] D. Kieras and D. Meyer, "An overview of the epic architecture for cognition and performance with application to human-computer interaction," *Human-Computer Interaction*, vol. 12, no. 4, 1997.

[137] G. Rizzolatti, L. Fogassi, and V. Gallese, "Parietal cortex: from sight to action," *Current Opinion in Neurobiology*, vol. 7, pp. 562–567, 1997.

[138] G. Rizzolatti, L. Fadiga, L. Fogassi, and V. Gallese, "The space around us," *Science*, pp. 190–191, 1997.

[139] P. Langley, "An cognitive architectures and the construction of intelligent agents," in *Proceedings of the AAAI-2004 Workshop on Intelligent Agent Architectures*, Stanford, CA., 2004, p. 82.

[140] D. Choi, M. Kaufman, P. Langley, N. Nejati, and D. Shapiro, "An architecture for persistent reactive behavior," in *Third International Joint Conference on Autonomous Agents and Multi-Agent Systems*. New York: ACM Press, 2004, pp. 988–995.

[141] P. Langley, "Cognitive architectures and general intelligent systems," *AI Magazine*, 2006, in Press.

[142] D. Benjamin, D. Lyons, and D. Lonsdale, "Adapt: A cognitive architecture for robotics," in *2004 International Conference on Cognitive Modeling*, A. R. Hanson and E. M. Riseman, Eds., Pittsburgh, PA, July 2004.

[143] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. RA-2, no. 1, pp. 14–23, 1986.

[144] M. P. Shanahan, "A cognitive architecture that combines internal simulation with a global workspace," *Consciousness and Cognition*, 2006, to Appear.

[145] ——, "Emotion, and imagination: A brain-inspired architecture for cognitive robotics," in *Proceedings AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness*, 2005, pp. 26–35.

[146] ——, "Cognition, action selection, and inner rehearsal," in *Proceedings IJCAI Workshop on Modelling Natural Action Selection*, 2005, pp. 92–99.

[147] B. J. Baars, *A Cognitive Theory of Consciousness*. Cambridge University Press, 1998.

[148] ——, "The conscious assess hypothesis: origins and recent evidence," *Trends in Cognitive Science*, vol. 6, no. 1, pp. 47–52, 2002.

[149] I. Aleksander, "Neural systems engineering: towards a unified design discipline?" *Computing and Control Engineering Journal*, vol. 1, no. 6, pp. 259–265, 1990.

[150] O. Michel, "Webots: professional mobile robot simulation," *International Journal of Advanced Robotics Systems*, vol. 1, no. 1, pp. 39–42, 2004.

[151] J. Weng, "Developmental robotics: Theory and experiments," *International Journal of Humanoid Robotics*, vol. 1, no. 2, pp. 199–236, 2004.

[152] ——, "A theory of developmental architecture," in *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, La Jolla, October 2004.

[153] ——, "A theory for mentally developing robots," in *Proceedings of the 2nd International Conference on Development and Learning (ICDL 2002)*. IEEE Computer Society, 131–140 2002.

[154] J. Weng, W. Hwang, Y. Zhang, C. Yang, and R. Smith, "Developmental humanoids: Humanoids that develop skills automatically," in *Proceedings the first IEEE-RAS International Conference on Humanoid Robots*, Cambridge, MA, 2000.

[155] J. Weng and Y. Zhang, "Developmental robots - a new paradigm," in *Proc. Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, 2002.

[156] J. L. Krichmar and G. M. Edelman, "Brain-based devices for the study of nervous systems and the development of intelligent machines," *Artificial Life*, vol. 11, pp. 63–77, 2005.

[157] J. L. Krichmar, D. A. Nitz, J. A. Gally, and G. M. Edelman, "Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task," *Proceedings of the National Academy of Science, USA*, vol. 102, pp. 2111–2116, 2005.

[158] J. L. Krichmar, A. K. Seth, D. A. Nitz, J. G. Fleisher, and G. M. Edelman, "Spatial navigation and causal analysis in a brain-based device modelling cortical-hippocampal interactions," *Neuroinformatics*, vol. 3, pp. 197–221, 2005.

[159] J. L. Krichmar and G. N. Reeke, "The darwin brain-based automata: Synthetic neural models and real-world devices," in *Modelling in the neurosciences: from biological systems to neuromimetic robotics*, G. N. Reeke, R. R. Poznanski, K. A. Lindsay, J. R. Rosenberg, and O. Sporns, Eds. Boca Raton: Taylor and Francis, 2005, pp. 613–638.

[160] A. Seth, J. McKinstry, G. Edelman, , and J. L. Krichmar, "Active sensing of visual and tactile stimuli by brain-based devices," *International Journal of Robotics and Automation*, vol. 19, no. 4, pp. 222–238, 2004.

[161] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *Journal of Neurscience*, vol. 2, no. 1, pp. 32–48, 1982.

[162] C. Burghart, R. Mikut, R. Stiefelhagen, T. Asfour, H. Holzapfel, P. Steinhaus, and R. Dillman, "A cognitive architecture for a humanoid robot: A first approach," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2005)*, 2005, pp. 357–362.

[163] I. Horswill, "Tagged behavior-based systems: Integrating cognition with embodied activity," *IEEE Intelligent Systems*, pp. 30–38, 2001.

[164] ——, "Cerebus: A higher-order behavior-based system," *AI Magazine*, 2006, in Press.

[165] A. Arkin, *Behavior-based Robotics*. Cambridge, MA: MIT Press, 1998.

[166] R. A. Brooks, C. Breazeal, M. Marajanovic, B. Scassellati, and M. M. Williamson, "The cog project: Building a humanoid robot," in *Computation for Metaphors, Analogy and Agends*, ser. Springer Lecture Notes in Artificial Intelligence, C. L. Nehaniv, Ed., vol. 1562. Berlin: Springer-Verlag, 1999.

[167] B. Scassellati, "Theory of mind for a humanoid robot," *Autonomous Robots*, vol. 12, pp. 13–24, 2002.

[168] A. M. Leslie, "Tomm, toby, and agency: Core architecture and domain specificity," in *Mapping the Mind: Specificity in Cognition and Culture*, L. A. Hirschfeld and S. A. Gelman, Eds. Cambridge, MA: Cambridge University Press, 1994, pp. 119–148.

[169] S. Baron-Cohen, *Mindblindness*. Cambridge, MA: MIT Press, 1995.

[170] C. Breazeal, *Sociable Machines: Expressive Social Exchange Between Humans and Robots*, ser. Unpublished Doctoral Dissertation. Cambridge, MA: MIT, 2000.

[171] ——, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, pp. 119–155, 2003.

[172] E. Thelen, "Time-scale dynamics and the development of embodied cognition," in *Mind as Motion – Explorations in the Dynamics of Cognition*, R. F. Port and T. van Gelder, Eds. Cambridge, Massachusetts: Bradford Books, MIT Press, 1995, pp. 69–100.

[173] C. von Hofsten, "An action perspective on motor development," *Trends in Cognitive Science*, vol. 8, pp. 266–272, 2004.

[174] ——, "On the development of perception and action," in *Handbook of Developmental Psychology*, J. Valsiner and K. J. Connolly, Eds. London: Sage, 2003, pp. 114–140.

[175] G. Sandini, G. Metta, and J. Konczak, "Human sensori-motor development and artificial systems," 1997.

[176] A. Billard, "Imitation," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 2002, pp. 566–569.

[177] R. Rao, A. Shon, and A. Meltzoff, "A bayesian model of imitation in infants and robots," in *Imitation and Social Learning in Robots, Humans, and Animals: Behaviour, Social and Communicative Dimensions*, K. Dautenhahn and C. Nehaniv, Eds. Cambridge University Press, 2004.

[178] K. Dautenhahn and A. Billard, "Studying robot social cognition within a developmental psychology framework," in *Proceedings of Eurobot*

99: Third European Workshop on Advanced Mobile Robots, Switzerland, 1999, pp. 187–194.

[179] A. N. Meltzoff and M. K. Moore, "Explaining facial imitation: A theoretical model," *Early Development and Parenting*, vol. 6, pp. 179–192, 1997.

[180] A. N. Meltzoff, "The elements of a developmental theory of imitation," in *The Imitative Mind: Development, Evolution, and Brain Bases*, A. N. Meltzoff and W. Prinz, Eds. Cambridge: Cambridge University Press, 2002, pp. 19–41.

[181] J. Nadel, C. Guerini, A. Peze, and C. Rivet, "The evolving nature of imitation as a format for communication," in *Imitation in Infancy*, J. Nadel and G. Butterworth, Eds. Cambridge: Cambridge University Press, 1999, pp. 209–234.

[182] G. S. Speidel, "Imitation: a bootstrap for learning to speak," in *The many faces of imitation in language learning*, G. E. Speidel and K. E. Nelson, Eds. Springer Verlag, 1989, pp. 151–180.

[183] C. Trevarthen, T. Kokkinaki, and G. A. Fiamenghi Jr., "What infants' imitations communicate: with mothers, with fathers and with peers," in *Imitation in Infancy*, J. Nadel and G. Butterworth, Eds. Cambridge: Cambridge University Press, 1999, pp. 61–124.

[184] B. Ogden, K. Dautenhahn, and P. Stribling, "Interactional structure applied to the identification and generation of visual interactive behaviour: Robots that (usually) follow the rules," in *Gesture and Sign Languages in Human-Computer Interaction*, ser. Lecture Notes LNAI, I. Wachsmuth and T. Sowa, Eds. Springer, 2002, vol. LNAI 2298, pp. 254–268.

[185] H. H. Clark, "Managing problems in speaking," *Speech Communication*, vol. 15, pp. 243–250, 1994.

[186] K. Doya, "What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?" *Neural Networks*, vol. 12, pp. 961–974, 1999.

[187] J. L. McClelland, B. L. NcNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory," *Psychological Review*, vol. 102, no. 3, pp. 419–457, 1995.

[188] N. P. Rougier, "Hippocampal auto-associative memory," in *International Joint Conference on Neural Networks*, 2001.

[189] A. Sloman and J. Chappell, "Altricial self-organising information-processing systems," in *International Workshop on the Grand Challenge in Non-classical Computation*, York, Apr. 2005.

[190] ——, "The altricial-precocial spectrum for robots," in *IJCAI '05 – 19th International Joint Conference on Artifi cial Intelligence*, Edinburgh, 30 July – 5 Aug. 2005.

[191] L. Craighero, M. Nascimben, and L. Fadiga, "Eye position affects orienting of visuospatial attention," *Current Biology*, vol. 14, pp. 331–333, 2004.

[192] L. Craighero, L. Fadiga, G. Rizzolatti, and C. A. Umilta, "Movement for perception: a motor-visual attentional effect," *Journal of Experimenal Psychology: Human Perception and Performance.*, 1999.

[193] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti, "Action recognition in the premotor cortex," *Brain*, vol. 119, pp. 593–609, 1996.

[194] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, pp. 131–141, 1996.

[195] E. J. Gibson and A. Pick, *An Ecological Approach to Perceptual Learning and Development*. Oxford University Press, 2000.

[196] L. Olsson, C. L. Nehaniv, and D. Polani, "From unknown sensors and actuators to actions grounded in sensorimotor perceptions," *Connection Science*, vol. 18, no. 2, 2006.

[197] T. Ziemke, "Are robots embodied?" in *Proceedings of the First International Workshop on Epigenetic Robotics — Modeling Cognitive Development in Robotic Systems*, ser. Lund University Cognitive Studies, Balkenius, Zlatev, Dautenhahn, Kozima, and Breazeal, Eds., vol. 85, Lund, Sweden, 2001, pp. 75–83.

[198] ——, "What's that thing called embodiment?" in *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, ser. Lund University Cognitive Studies, Alterman and Kirsh, Eds. Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 1134–1139.