Gruppo Nazionale di Bioingegneria, Università Degli Studi di Padova, XXV Scuola Annuale, Neuro-Robotica: Neuroscienze e Robotica per lo Sviluppo di Macchine Intelligenti, Bressanone, 25-29 settembre 2006

Cognition in Artificial Systems

David Vernon
LIRA-Lab, University of Genova, Italy
email: vernon@ieee.org

1 Introduction to Cognition

Cognition — being able to understand what's going on around you and being able to adapt and improvise accordingly — defies easy definition. It has been equated with rationality and reasoning, deliberation and abstract thought and, while such concerns are clearly relevant, it is by no means clear that they necessarily form the essence of cognition. It all comes down to what we mean when we say we understand something and what we deem to be an appropriate way to adapt our behaviour. Does a central heating thermostat understand the need for warmth (and the consequences of not getting it) when it detects that the temperature of a room is too cold and switches on the heating? Certainly not in any meaningful way.[1] It seems that such a scenario is too trivial to be interesting. *Understanding* becomes interesting only if there are many of factors to be considered in assessing a situation and, especially, if they are complicated: they might conflict, there may be some essential information missing, they might be constantly changing, or they might be simply incorrect.[2] These complications mean that it is not sufficient for a cognitive system just to react to present circumstances, to how things are, to make some best-possible choice based on some or all of the available information, and then proceed on.

Cognition implies an ability to understand how things might possibly be, not now but at some future time, and take this into consideration. Remembering what happened at some point in the past can help in anticipating future events, so memory is important too: using the past to predict the

[1] The idea that a thermostat could legitimately be viewed as thinking and having beliefs (specifically, that the room is too hot, too cold, or ok) is due to John McCarthy of Stanford University and appears in a 1983 paper 'The Little Thoughts of Thinking Machines' in *Psychology Today* [1]; see `http://www.cse.msu.edu/ cse841/papers/McCarthy.pdf`. McCarthy is often referred to as the father of Artificial Intelligence.

[2] John McCarthy did extend his thermostat scenario to more complex situation where 'compromise' would be required. His essential point was that it helps for people to think about machines as having the capacity to think both when using them and when designing them.

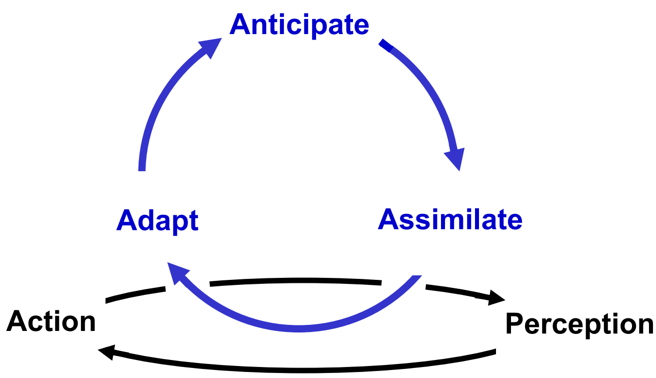

Figure 1: Breaking the 'here-and-now barrier': cognition as a cycle of anticipation (or prospection), assimilation, and adaptation, embedded in, contributing to, and benefitting from a continuous process of action and perception.

future[3] 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 (see Figure 1).

Cognition breaks through the 'here-and-now barrier' and takes us into the future with the help of the past, in a way that allows the system to adapt and improve.

But there are still some tough questions to be answered: what makes an action the right one to choose? Having broken through the here-and-now barrier, what type of behaviour does cognition enable? This opens up another dimension of the problem: what motivates cognition? And yet another pops into view: what makes cognition possible? Cognitive skills can improve, but what do you need to get started?

Unfortunately, no one yet knows how to design and build an artificial cognitive system. There is no shortage of ideas and many alternative approaches have been proposed, but a complete convincing artificial cognitive system hasn't yet been developed. Understanding cognition, and modelling, designing, and building artificial cognitive systems are challenging long-term research problems.

The goal of this article is to provide you with a short overview of the many topics involved in cognition and cognitive systems and to do so from several perspectives. This presents a challenge for three reasons. First, it is a big field. It embraces artificial intelligence, psychology, neuroscience, non-linear dynamical systems theory, synergetics, autonomous systems theory, machine learning, pattern recognition, computer vision, haptic sensing, aural perception, cybernetics, neural networks, epistemology, philosophy, linguis-

[3] Berthoz puts it very succinctly: 'Memory is used primarily to predict the consequences of future action by recalling those of past action' [2].

tics, semiotics, robotics, manipulation, communication, and many others. All these impact in some way or other on cognitive systems.

Second, it is not always easy to say where one should draw the boundary between cognition in natural systems and cognition in artificial computer-based systems. The boundary between these two fields is fuzzy. Sometimes, they form a symbiotic relationship, one learning from the other's successes and failures. Other times, they are antagonistic: limitations in our understanding of one field sometimes act as a brake on developments in the other field.

The third reason why the goal of the article is challenging is that there is no universal agreement as to what cognition is, in the first place! Our aim here is to make sense of all this.

In the limited space available, we can't hope to address all the topics mentioned above. However, we will try to identify the full scope of cognition and cognitive systems and we will try to provide a useful working definition, one that strikes a balance between being broad enough to do service to the many views that people have on cognition and deep enough to help in the formulation of theories and models. We will then present a summary of the many approaches that people adopt in researching and developing cognitive systems.

Ultimately, this article is intended to give you a clear understanding of the scope of the domain, its alternative approaches, and their underlying differences. Perhaps most important of all, it will give you a solid grasp of the issues that need to be addressed in striving for the goal of creating a true artificial cognitive system.

1.1 A Definition of Cognitive Systems

As we noted already, there is no universal consensus on what exactly cognition is. Nonetheless, to get things started, we will make an initial attempt to define the area of cognitive systems. We will expand on this later in the book.

A cognitive system produces effective — adaptive, anticipatory, and purposive goal-directed — 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 should have some degree of plasticity and be resilient in the face of the unexpected. The characteristic of anticipation and prospective behaviour is also crucial as it allows the system to operate across a variety of time-scales, in the here-and-now, but extending into the future. Thus, a cognitive system is capable of more than reactive stimulus-response behaviour (which might be quite complex in its own right).

Some authors in discussing the development of cognitive systems go even further than this. For example, Brachman [3] 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 current approach to carrying out a task or to know when it needed new information in order to complete it. Hollnagel [4] 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 self-development.

1.2 Emulation or Simulation?

Before going any further, we need to be clear exactly what we are trying to do in attempting to develop an artificial cognitive system. Are we trying to develop an artifact that emulates the cognitive behaviour and capabilities of human beings, or are we trying to simulate the actual process by which a human being effects such behaviour and capabilities. The distinction, which is really the re-appearance of the fuzzy boundary between biological and computer cognition, is important because in the case of emulation it is only the end product — the system behaviour — that is crucial. On the other hand, in the case of simulation, it is necessary to be as faithful as possible to the human or biological process which underpin cognition. This doesn't mean that in the case of emulation we should ignore completely the biological systems. On the contrary, we need to look somewhere for inspiration in trying to deal with this very complex topic and biological systems are the only exemplars of cognition we have. Clearly, we should draw as much inspiration as possible from them and from what is known about biological cognition. In this article, we assume that our task is to emulate cognitive capability. That is, we want to create artificial cognitive systems — systems with all the desirable attributes set out above — but we don't necessarily want to make any strong claims that the resultant models are either biologically plausible or that they are viable models of human cognition. That said, we should look to the constraints of biological plausibility to provide some guidance in our search for a model of artificial cognition.

2 A Survey of Cognition Paradigms

There are several quite distinct approaches to understanding and synthesis of cognitive systems, including physical symbol systems, connectionism, artificial life, dynamical systems, and enactive systems[5, 6]. Each of these makes significantly different assumptions about the nature of cognition, its purpose, and the manner in which cognition is achieved. 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, and based to a lesser or greater extent on principles of self-organization.

2.1 Cognitivist Models

Cognitivism asserts that cognition involves computations defined over symbolic representations, in a process whereby information about the world is abstracted by perception, represented using some appropriate symbol set, reasoned about, and then used to plan and act in the world. This approach has also been labelled by many as the *information processing* approach to cognition [7, 8, 9, 10, 6, 11, 12]. The discipline of cognitive science is often (erroneously) identified exclusively with this particular approach [12]. It is, however, by no means the only paradigm in cognitive science and there are indications that the discipline is migrating away from its stronger interpretations [5].

In most cognitivist approaches concerned with the creation of artificial cognitive systems, the symbolic representations are the 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 vision systems: these designer-dependent representations are the idealized descriptions of a human cognitive entity and, as such, they effectively bias the system (or 'blind' it [13]) and constrain it to an domain of discourse that is dependent on and, a consequence of, the cognitive effects of human activity. This approach works well 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' [14] between perception and possible interpretation, a gap that is normally plugged by embedding programmer knowledge or enforcing expectation-driven constraints [15] to render a system practicable in a given space of problems.

2.2 Emergent Systems

Emergent systems, embracing connectionist, dynamical, and enactive systems, take a very different view of cognition. Here, cognition is a process of self-organization whereby the system is continually re-constituting itself in real-time to maintain its operational identity through moderation of mutual system-environment interactions and co-determination [16]. 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. This has significant implications for the nature of perception and cognitive vision. 'Perceiving is not strictly speaking in the animal or an achievement of the animal's nervous system, but rather is a process in an animal-environment system' [12]. Co-determination is one of the key differences between the emergent paradigm and the cognitivist paradigm, wherein an objective reality common to all cognitive agents is assumed. For emergent systems, vision provides appropriate sensory data to enable effective action [16] but it does so as a consequence of the system's actions. In the emergent paradigm, cognitive vision is functionally dependent on the richness of the action interface [17].

2.2.1 Connectionist Models

One of the original motivations for work on emergent systems was disaffection with the sequential, atemporal, and localized character of symbol-manipulation based cognitivism [6]. Emergent systems, on the other hand, depend on parallel, real-time, and distributed architectures. 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' [6]. 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" [6]. Indeed, the 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.

Connectionist approaches are for the most part associative learning systems in which the learning phase is either unsupervised (self-organizing) or supervised (trained). For example, hand-eye coordination can be learned by a Kohonen neural network from the association of proprioceptive and exteroceptive stimuli [18, 19]. 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) [20]. An example of a connectionist system which exploits the co-dependency of perception and action in a developmental setting can be found in [21]. This is a biologically-motivated connectionist 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 [22].

2.2.2 Dynamical Models

Dynamical systems theory has been used to complement classical approaches in artificial intelligence [23] and it has also been deployed to model natural and artificial cognitive systems [12, 11, 24]. 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 [12].

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.

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 [25, 26, 12, 27, 28], the principled feasibility of higher-order cognitive faculties has yet to be validated.

Dynamical approaches differ from connectionist systems in a number of ways [12, 11, 24]. Suffice it here to note that the connectionist system is often defined by a general differential equation which is actually a schema that defines the operation of many (neural) units. That is, the differential equation applies to each unit and each unit is just a replication of a common type. This also means that there will be many independent state variables, one for each unit. Dynamical systems, on the other hand, are not made up of individual units all having the same defining equation and can't typically be so decomposed. Typically, there will be a small number of state variables that describe the behaviour of the system as a whole.

2.2.3 Enactive Systems Models

Cognitivism, by definition, involves a view of cognition that requires the representation of a given objective pre-determined world [6, 24]. Enaction [29, 30, 31, 32, 6, 13, 16] adopts a fundamentally different stance: cognition is a process whereby the issues that are important for the continued existence of the cognitive entity are 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 representations. Instead there is an enactive interpretation: a context-based choosing of relevance. Is the role of cognition to abstract objective structure and meaning through perception and reasoning? Or, is it to uncover unspecified regularity and order that can then be construed as meaningful because they facilitate the continuing operation and evolution of the cognitive system?

Enaction adopts the second stance, one that is actually more neutral, 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 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 and its space of interaction with the environment.

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 [29, 30, 31, 32, 6, 13, 16].

The enactive approach is mirrored in the ideas of self-maintenant system and recursive self-maintenant systems [33]. Here autonomy is defined 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. Self-maintenant systems make active contributions to their own persistence but do not contribute to the maintenance of the conditions for persistence. Conversely, recursive self-maintenant systems do contribute actively to the conditions for persistence and can deploy different processes of self-maintenance depending on environmental conditions.

2.3 Hybrid Models

Considerable effort has gone into developing approaches which combine aspects of the emergent systems and cognitivist systems [17, 34, 35]. These hybrid approaches have their roots in strong criticism of the use of explicit programmer-based knowledge in the creation of artificially-intelligent systems [36] and in the development of active 'animate' perceptual systems [37] 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 [17]. 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. Action precedes perception and 'cognitive systems need to acquire information about the external world through learning or association' [34]. Hybrid systems are in many ways consistent with emergent systems while still exploiting programmer-centred (but not programmer-populated) representations (for example, see [38]).

2.4 Which Approach is Right?

Each approach has its own strengths and weaknesses, and its proponents and critics. 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. It has been argued [39] 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), 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 [39]. 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 [39] and has so far been employed more as an analysis tool than as a tool for the design and synthesis of cognitive systems [40, 39]. The extent to which this will change, and the speed with which it will do so, is uncertain. Some people feel that hybrid approaches offer the best of both worlds but it is unclear how well one can combine what are ultimately highly antagonistic underlying philosophies. Opinion is divided, with arguments both for (*e.g.* [5, 41, 42]) and against (*e.g.* [39]).

3 Cognitive Architectures

3.1 What is a Cognitive Architecture?

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 [43, 44, 45]. Since cognitive architectures represent the fixed part of cognition, they cannot accomplish anything in their own right and need to be provided with knowledge to peform any given task. This combination of a given cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model*. Note that in most cognitivist systems the knowledge incorporated into the model is determined by the (human) modeller. The specification of a cognitive architecture consists of its representational assumptions, the characteristics of its memories, and the processes that operate on those memories.

Although used freely by proponents of the cognitivist, emergent, and hybrid approaches to cognitive systems, the term cognitive architecture originated in fact with the seminal work of Newell *et al.* [46, 47, 48]. Consequently, the term has a very specific meaning in the cognitivist paradigm in that cognitive architectures represent attempts to create unified theories of cognition [49, 47, 50]. Because unified theories of cognition are concerned with the computational understanding of human cognition, cognitive architectures too are, from this perspective, concerned with human cognition [47][4] On the other hand, for emergent approaches to cognition that have a focus on development from a primitive state to a fully cognitive state over the life-time of the system, the architecture of the system is equivalent to its phylogenic configuration: the innate capabilities with which it is endowed at the beginning of its life-time and which don't have to be learned (but may be developed further).

3.2 A Short Survey of Cognitive Architectures

The Soar system [52, 48, 53, 54] is a typical example of a cognitivist architecture. It is a production (or rule-based) system that operates in a cyclic manner, with a production cycle and a decision cycle. If the decision cycle leads to an impasse, Soar resolves it by creating sub-goal and then encapsulating in a new production rule. This is the only form of learning that occurs in Soar.

EPIC [55] is a cognitive architecture that was designed to link high-fidelity models of perception and motor mechanisms with a production sys-

[4]Even though some [51] have argued that the term cognitive architecture should be reserved for systems that model human cognition, suggesting that the term "agent architecture" as a better term to refer to general intelligent behaviour, including human cognition, it has become common-place to use cognitive architecture in this more general sense, both in the cognitivist and emergent paradigms.

tem. 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. EPIC does not have any learning mechanism.

Anderson's ACT-R (Adaptive Control of Thought - Rational) [56, 50] cognitive architecture 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. 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. 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 [49, 50]. That said, the perceptuo-motor system in ACT-R is based on the EPIC architecture [55] 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 [56]. Anderson *et al.* recognize that this is a short-coming, 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.

Langley's ICARUS cognitive architecture [57, 45, 58, 59] 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, operating according to the conventional recognize-act cycle. 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.

Burghart *et al.* [60] present a cognitive architecture (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.

Some authors, e.g. Benjamin *et al.* [61], argue that existing cognitive architectures such as Soar, ACT-R, and EPIC, don't easily support certain mainstream robotics paradigms such as adaptive dynamics and active perception. 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.

Horswill [62, 63] also 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. He proposes a 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 [64] and Arkin [65]) 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.

Shanahan [66, 67, 68, 69] 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.

Autonomous agent robotics (AAR) and behaviour-based systems represents another 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 [64]. On the other hand, Christensen and Hooker [39] argue that

AAR is not sufficient either as a principled foundation for a general theory of situated cognition. and they propose a new interactivist-constructivist (I-C) approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [70]. 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.

Weng [71, 72, 73] introduced a cognitive architecture that is specifically focussed on the issue of development by which he means that the processing accomplished by the architecture is no 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.

4 Implications for the Development of Cognition in Artificial Systems

We will conclude by drawing together the main issues raised in the foregoing and summarize some of the key features that an artificial cognitive system should probably exhibit, in particular those that adhere to a developmental approach.

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 representation or encoding but together achieving the cognitive goal of effective behaviour. This network forms the system's phylogenetic configuration and its innate abilities.

Second, a developmental cognitive architecture must be capable of some form of self-modification, not simply by adjusting system parameters (such as when using machine learning to effect parameter estimation) but by facilitating the modification of the very structure and organization of the system itself so that it is capable of changing its system dynamics, expanding its repertoire of actions, and thereby adapting to new circumstances. This development should be driven by both explorative and social motives, one 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 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 [50] or implicitly like Shanahan's global workspace dynamical architecture [66] — and use this to modulate the actual behaviour of the system.

Fourth, 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, if the epistemological understanding of the developed systems is required to be consistent with that of other cognitive agents such as humans. 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 necessary and sufficient for artificial cognitive systems remains an open question.

We have touched in this article on some of the main topics underpinning artificial cognitive systems. But there are many other issues that need to be addressed and about which we've said very little. These include the tight relationship between cognition, action, and perception. There is also the matter of autonomy: whether or not cognition necessarily implies autonomy and, if it does, what are the implications. Similarly, cognition is tied up the the issue of embodiment, so one has to consider the different forms of embodiment that are possible, and the role that embodiment plays in the cognitive process. The issue of development leads to the ontogenesis of cognitive systems and the trade-off between phylogeny and ontogeny in the altricial/precocial spectrum. There are many other facets to cognition too, each of which need to be considered. These include deliberation, reasoning, development, learning, representation, memory, language, communication, goal achievement, motivation, prospection, planning, recognition, and concept formation. Finally, any complete treatment of cognition should also consider the role (or relevance) of consciousness, as well as its philosophical, psychological, and neuroscientific foundations of cognition. We leave these issue to another article.

Acknowledgments

This work was funded by the European Commission as part of *RobotCub*, ROBotic Open-architecture Technology for Cognition, Understanding and Behaviors, project 004370, `www.robotcub.eu`.

References

[1] J. McCarthy. The little thoughts of thinking machines. *Psychology Today*, 17(12):46–49, 1983.

[2] A. Berthoz. *The Brain's Sense of Movement.* Harvard University Press, Cambridge, MA, 2000.

[3] R. J. Brachman. Systems that know what they're doing. *IEEE Intelligent Systems*, 17(6):67–71, December 2002.

[4] E. Hollnagel and D. D. Woods. Cognitive systems engineering: New wind in new bottles. *International Journal of Human-Computer Studies*, 51:339–356, 1999.

[5] A. Clark. *Mindware – An Introduction to the Philosophy of Cognitive Science.* Oxford University Press, New York, 2001.

[6] F. J. Varela. Whence perceptual meaning? A cartography of current ideas. In F. J. Varela and J.-P. Dupuy, editors, *Understanding Origins – Contemporary Views on the Origin of Life, Mind and Society*, Boston Studies in the Philosophy of Science, pages 235–263. Kluwer Academic Publishers, 1992.

[7] D. Marr. Artificial intelligence – A personal view. *Artificial Intelligence*, 9:37–48, 1977.

[8] J. Haugland. Semantic engines: An introduction to mind design. In J. Haugland, editor, *Mind Design: Philosophy, Psychology, Artificial Intelligence*, pages 1–34, Cambridge, Massachusetts, 1982. Bradford Books, MIT Press.

[9] S. Pinker. Visual cognition: An introduction. *Cognition*, 18:1–63, 1984.

[10] J. F. Kihlstrom. The cognitive unconscious. *Science*, 237:1445–1452, September 1987.

[11] E. Thelen and L. B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action.* MIT Press / Bradford Books Series in Cognitive Psychology. MIT Press, Cambridge, Massachusetts, 1994.

[12] J. A. S. Kelso. *Dynamic Patterns – The Self-Organization of Brain and Behaviour.* MIT Press, 3rd edition, 1995.

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

[14] 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*, 22(12):1349–1380, December 2000.

[15] J. Pauli and G. Sommer. Perceptual organization with image formation compatibilities. *Pattern Recognition Letters*, 23(7):803–817, 2002.

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

[17] G. H. Granlund. The complexity of vision. *Signal Processing*, 74:101–126, 1999.

[18] M. Jones and D. Vernon. Using neural networks to learn hand-eye coordination. *Neural Computing and Applications*, 2(1):2–12, 1994.

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

[20] R. Linsker. Self-organization in a perceptual network. *Computer*, pages 105–117, March 1988.

[21] G. Metta, G. Sandini, and J. Konczak. A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12(10):1413–1427, 1999.

[22] H. Buxton and A. J. Howell. Active Vision Techniques for Visually Mediated Interaction. In *International Conference on Pattern recognition*, Quebec City, Canada, August 2002.

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

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

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

[26] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, 1979.

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

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

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

[30] H. Maturana. The organization of the living: a theory of the living organization. *Int. Journal of Man-Machine Studies*, 7(3):313–332, 1975.

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

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

[33] M. H. Bickhard. Autonomy, function, and representation. *Artificial Intelligence, Special Issue on Communication and Cognition*, 17(3-4):111–131, 2000.

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

[35] G. H. Granlund. Cognitive vision – background and research issues. Research report, Linköping University, 2002.

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

[37] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[38] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe. A boosted particle filter: Multitarget detection and tracking. In T. Pajdla and J. Matas, editors, *Proceeding of the 8th European Conference on Computer Vision, ECCV 2004*, volume 3021 of *LNCS*, pages 28–39. Springer-Verlag, 2004.

[39] 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.

[40] 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.

[41] G. Granlund. Organization of architectures for cognitive vision systems. In H. I. Christensen and H.-H. Nagel, editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, LNCS, pages 39–58, Heidelberg, 2005. Springer-Verlag.

[42] J. P. Crutchfield. Dynamical embodiment of computation in cognitive processes. *Behavioural and Brain Sciences*, 21(5):635–637, 1998.

[43] 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*, 12:301–309, 1997.

[44] 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*, 55:1–14, 2001.

[45] P. Langley. An adaptive architecture for physical agents. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 18–25, Compiegne, France, 2005. IEEE Computer Society Press.

[46] A. Newell. The knowledge level. *Artificial Intelligence*, 18(1):87–127, March 1982.

[47] A. Newell. *Unified Theories of Cognition.* Harvard University Press, Cambridge MA, 1990.

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

[49] M. D. Byrne. Cognitive architecture. In J. Jacko and A. Sears, editors, *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*, pages 97–117. Lawrence Erlbaum, Mahwah, NJ, 2003.

[50] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060, 2004.

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

[52] J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: an architecture for general intelligence. *Artificial Intelligence*, 33(1–64), 1987.

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

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

[55] D. Kieras and D. Meyer. An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 1997.

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

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

[58] 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*, pages 988–995, New York, 2004. ACM Press.

[59] P. Langley. Cognitive architectures and general intelligent systems. *AI Magazine*, 2006. In Press.

[60] 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)*, pages 357–362, 2005.

[61] D.P. Benjamin, D. Lyons, and D. Lonsdale. Adapt: A cognitive architecture for robotics. In A. R. Hanson and E. M. Riseman, editors, *2004 International Conference on Cognitive Modeling*, Pittsburgh, PA, July 2004.

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

[63] I. Horswill. Cerebus: A higher-order behavior-based system. *AI Magazine*, 2006. In Press.

[64] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2(1):14–23, 1986.

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

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

[67] M. P. Shanahan and B. Baars. Applying global workspace theory to the frame problem. *Cognition*, 98(2):157–176, 2005.

[68] M. P. Shanahan. Emotion, and imagination: A brain-inspired architecture for cognitive robotics. In *Proceedings AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness*, pages 26–35, 2005.

[69] M. P. Shanahan. Cognition, action selection, and inner rehearsal. In *Proceedings IJCAI Workshop on Modelling Natural Action Selection*, pages 92–99, 2005.

[70] W. D. Christensen and C. A. Hooker. An interactivist-constructivist approach to intelligence: self-directed anticipative learning. *Philosophical Psychology*, 13(1):5–45, 2000.

[71] J. Weng. Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, 1(2):199–236, 2004.

[72] J. Weng. A theory of developmental architecture. In *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, La Jolla, October 2004.

[73] J. Weng. 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.