Designed by Complexity: Adaptive General Intelligence Without Algorithms

Christopher Larcombe
University of Sussex, Brighton, BN1 9QH, UK.
E-mail: cl249@sussex.ac.uk

Spring 2009

Abstract

The open-ended acquisition of new skills and integration with existing skills is considered a necessary property of adaptive general intelligence. An argument is made that in order for it to be understood and replicated in an artificial system, an atypical level of description is necessary. Steps are taken to establish a more abstract perspective of cognitive phenomena, acknowledging the role of the observer. Cognitive architectures with fixed processing structures are criticised. It is concluded that truly autonomous adaptive general intelligence may only be possible in model artificial systems equivalent to self-organising structures 'programmed by nature'.

1 Introduction

This paper is concerned with the problem of open-ended skill acquisition in artificial systems, as a key step towards artificial intelligence with greater adaptivity, generality and flexibility. It is argued that the level of description commonly used when considering living systems, particularly humans, is inappropriate for dealing with this ability, and that a more abstract perspective is necessary.

An attempt is made to make several steps towards a new perspective, with the intention of re-framing the problem of integrating cognitive abilities entirely. The fixed architectures, algorithms, procesing structures and statistic frameworks characteristic of attempts to integrate multiple cognitive phenomenon in a single architecture are crtiticied as being fundamentally different to the living systems they attempt to model.

It is hoped that re-framing the problem will avoid such issues, allowing the design of artificial systems to more closely resemble living systems. As a further possible advantage, it is hoped that translating different cognitive phenomena into the perspective established will lead to new questions and conclusions.

1.1 Narrow Artificial Intelligence

Since the field of artificial intelligence (AI) was founded as an academic discipline in 1956, the dream of building a 'thinking' machine has been subject to heavy criticism, both in terms of its plausibility and achievability. Now approaching the second decade of the 21st century, over fifty years later, AI as it was originally envisaged has not yet emerged.

> "In from three to eight years we will have a machine with the general intelligence of an average human being."
> – M. Minsky (1970).

Despite the obvious far-reaching scientific, philosophical and technological implications of achieving such a goal, research into this form of AI, 'the thinking machine', has not been commonplace. Whether due to past overoptimistic claims and predictions, lack of ambition, computational resources or funding, so far the focus in AI has has been on more narrow, less ambitious applications. This is where AI has found its success. Narrow applications have grown into whole sub-fields within AI, such as computer vision, machine learning, data mining, natural language processing, knowledge representation and planning. In fact, fields such as data-mining have evolved so much so that they are more often given the label 'computer science', rather than artificial intelligence.

Computers can now successfully recognise continuous speech, categorise objects using vision techniques, play chess, plan time-tables, write music and even design electronic circuits. It is well accepted that computers can outperform our ability to do mental arithmetic, but recently more 'natural' cognitive abilities such as object recognition in computers are beginning to rival human being ability (Serre et al. 2007). Albeit partially inadvertently, we appear to be acquiring the components of a thinking machine, but those components remain disconnect from one another, in separate systems, in separate research departments, as separate pieces of software and hardware. How might we begin to integrate these components?

1.2 Towards Artificial General Intelligence

In the 1990s, a series of papers led to paradigm shift: embodiment, the idea that cognition and intelligence are determined by and largely depend on a physical body in the real-world (Brooks et al. 1998). The idea is powerful, multifaceted, and hard to define; the term is now used so widely it has even been described as "semantically vacuous" (Metzinger 2007). Put simply though, embodiment was the realisation that seemingly intelligent and adaptive behaviour could emerge from a purely physical system, without 'mental' properties or a central locus of control.

Taking inspiration from biology and early cybernetics work (Ashby 1960) it was hoped that with an embodied 'bottom-up' methodology more life-like robots could be constructed. Brooks (1991*a*,*b*) pioneered the field of behaviour-based robotics, which did indeed yield more life-like robotics, though the approach was eventually criticised for difficulties in 'scaling-up' beyond highly reactive, insect-like robots (Kirsh 1991, Brooks 1996).

More recently new fields have emerged, such as developmental robotics (Lungarella 2007, Lungarella et al. 2003), also known as epigenetic robotics, or ontogenic robotics. Advocating an embodied perspective and drawing on ideas from developmental psychology, neuroscience and dynamical systems theory, the emphasis is on modelling "the development of increasingly complex cognitive processes in natural and artificial systems and to understand how such processes emerge through physical and social interaction". While not openly and directly focusing on the development of general intelligence, a number of researchers have discussed possible steps in this direction (Smith & Breazeal 2007, Weng et al. 2001, Prince et al. 2005). It is the developmental approach that will be advocated in this paper.

In 2008, the first conference for 'artificial general intelligence' was held in Memphis, USA (Wang et al. 2008), directly focused on "the original and ultimate goal of AI – to create intelligence as a whole, by exploring all available paths, including theoretical and experimental computer science, cognitive science, neuroscience, and innovative interdisciplinary methodologies.".

It seems that now may be the time to start re-addressing issues of general intelligence, but where should we begin? One key problem identified by Prince et al. (2005) as a 'core concept in epigenetic robotics' is that of *ongoing emergence*: 'the continuous development and integration of new skills'. The authors state that ongoing emergence is exhibited when six different criteria are satisfied. One of which, "incorporation of new skills with existing skills", appears to be particularly absent from the research reviewed by the authors. Outside of the epigenetic robotics community, Sloman (2008) refers to the problem of integrating different capabilities in artificial systems as a problem of "scaling out" rather than "scaling up', commenting that the former does not relate to computational complexity, while scaling up a system does. By scaling out humans are able to use their capabilities creatively and flexibly in an integrated manor to solve problems; it is this ability that current artificial intelligence lacks.

For the concerns of this paper, open-ended skill acquisition will be considered a necessary step towards more general artificial intelligence that can 'scale out'. Open-ended development of useful behaviour i.e., learning new skills, maintaining them and incorporating them with previously acquired skills for subsequent use is a prerequisite or at least an essential component of any general intelligence system. It should also be noted that this is not a uniquely human ability, for example, chimpanzees are capable of learning and displaying a large variety of skills in an open-ended manor.

2 Architectures, Algorithms and Representations

In this section types of architecture, algorithms and representations will be discussed in the context of software for embodied agent control, applicable to real world agents (robots) or virtual agents (avatars). The word architecture is used broadly in multiple contexts, but here it is used to refer to the large-scale or high-level functional decomposition of the control system, as might be used by a software engineer or analyst.

As previously mentioned, in order to gain some level of generality in our agents we must investigate architectures that allow the integration of multiple components or skills. Choosing an appropriate architecture has been noted to be closely related to "defining an ontology for mental objects, states and processes (percepts, beliefs, desires, attitudes, intentions, moods, emotions, character, inferences, learning, etc)" (Sloman 1999). While the author is writing about architectures rather generally, it will be argued here that thinking in these terms is counter-productive when dealing with the specifics of open-ended skill acquisition and general intelligence. The author notes that "our ideas regarding the ontology to be supported by such an architecture are still very primative", but perhaps most importantly, for the concerns of this paper, states that "we should not assume an agent has a fixed architecture: part of the processes of learning and development may include changes ...". How does this reflect on our ontology?

In a recent review, Vernon et al. (2007) compared 14 state-of-the-art cognitive architectures in a survey, split across three different categories: cognitivist, emergent and hybrid[1]. The former two categories approximately correspond to two different paradigms, or perspectives on cognition (Vernon et al. 2007, pp 153). The survey scores each architecture for several different key 'characteristics of cognitive systems' including embodiment, perception, action, anticipation, adaptation, motivation and autonomy. Both SOAR and ACT-R, in addition to three other cognitivist architectures fail to address motivation, autonomy and only 'weakly address' adaptation. Furthermore, only one of which is embodied. Despite this, both SOAR and ACT-R were considered candidates for a Unified Theory of Cognition (Newell 1994), albeit not by the authors of the review paper. In comparison, the architectures based on the emergent paradigm were all embodied, all strongly address autonomy, two of which strongly addressed adaptation. Without autonomy, adaptation or motivation, it is difficult to imagine how an artificial system could ever be capable of open-ended skill acquisition.

What is the key difference evident in architecture design between the cognitivist and emergent paradigms? It is difficult to identify precisely, though perhaps the most striking difference is that cognitivist architectures often feature more "externally derived domain knowledge and processing structures", such as machine learning algorithms, statistic frameworks, discreet data-structures, modules or fixed high-level structure within the overall architecture. By using fixed data or processing structures limitations are imposed on the possible states the system can take during its operation, since the flexibility, learning potential and future development of such a

[1]a hybrid architecture features both cognitivist and emergent characteristics

system is determined and *constrained* by its state-space, i.e., the space of all possible states the system can be in at any one time.

For *narrow* applications such as industrial robotics, this may be an advantage rather than a problem, allowing precise, reliable behaviour and tractable learning. However, for more general open-ended applications, such as continuous skill acquisition, a different approach may be necessary. Whether or not a state-space is fixed does not imply the impossibility of learning or development, but rather restricts what is possible to certain emergent domains, which the designer may fail to anticipate. The key point is that the greater the number of *fixed* "processing structures", the more restricted the state-space will be.

Why do we impose these limitations on our systems? It seems that at least part of the problem originates from our assumed ontology, or *perspective on what we are modelling.* By assuming the inherent existence and separability of different cognitive functions, both high and low level, modularity and structure is imposed in the designed model, i.e., the assumptions are projected onto the design, within the architecture or its constituents. By acknowledging the observer-dependence and relativistic existence of cognitive phenomena such as perception, action, linguistic ability, as well as knowledge, memory, and other skills, a new framework can be constructed, more suited to the problem of understanding the integration of these apparently separate functions. A perspective without discreet, indivisible components and processes.

To further elucidate why this is a problem, we need to reconsider what we are modelling. By changing to a more abstract perspective and analysing how structures change over time, we can see how different these cognitive architectures are to the real systems we are trying to emulate.

3 A Change in Perspective

By changing the way we look at the world around us and ourselves within it, we can re-frame the problem entirely. In this section concepts from dynamical systems theory and complexity science are used to emphasise the observer-dependence of systems and how they change over time. It is hoped that doing so will allow us to design artificial systems that more closely resemble living systems. As a further possible advantage, it is hoped that translating different cognitive phenomena into the new perspective will lead to new questions and conclusions.

3.1 Complexity and the Observer

Complex systems theory, or complexity theory, can be defined as the inter-disciplinary study of the common properties of dynamically changing systems, stating that:

> Critically interacting components self-organise to form potentially evolving structures exhibiting a hierarchy of emergent system properties.

That is, under certain conditions hierarchical organisation can emerge in systems containing many interacting components, and those components may themselves constitute larger components. Complexity is concerned with describing the way in which components in such systems change over time in relation to other components, rather than describing the properties of the components themselves in isolation.

Hierarchies are in fact ubiquitous in nature, and may also be referred to as 'levels of description'. For example, a human being is not just a person, but a collection of organs, *or* a collection of cells, *or* a collection of molecules, etc. Each are equally valid and equally true. Such 'clusters' of parts are also evident on larger scales, such as cities, planets, solar systems, galaxies, and superclusters of galaxies. The theme is this: each whole contains parts, which are themselves constituted by smaller parts. The level of description used can only be justified by how *useful* it is. A mechanic is not concerned with the molecular make-up of the components of a car, nor the number of cars on the road and their geographic distribution. In a similar sense it is argued here that looking at a human being as a *person* with beliefs, desires, and intentions, as well other living systems is counter-productive for engineering purposes. In order to construct *adaptive* artificial systems that more closely resemble living systems a different level of description is necessary.

In dynamical systems theory, models are constructed in order to describe how *parts of a system* interact. Such models usually take the form of differential equations, consisting of variables and parameters. Variables describe something that can change, while parameters describe something that does not, but still has an affect. For example, consider the following two simple differential equations:

$$\frac{dy}{dt} = -\frac{1}{2}y + p \qquad \frac{dp}{dt} = \frac{1}{2000} \tag{1}$$

where y and p are *variables* which vary continuously over time, t. Variables in differential equations can potentially represent any continuously changing observable phenomena at any spatial or temporal scale, for example, the speed of a car, the rotation of a planet, the activity of a neuron, share values in the stock market. Physics, for example, is essentially the practice of discovering or creating equations which can be applied to the real world in this same sense. It is the way in which the variables (and parameters) in an equation relate to each other that determine how accurately the model describes reality. Of course, capturing the 'essence' of a changing system is not always an easy task, and only certain models will describe real phenomena.

In dynamical systems terminology, something which can be given a value that changes over time is referred to as a *collective variable*. Similarly, in synergetics (Haken 1988) this is referred to as an *order parameter*. Both are *emergent* in the sense that such variables only exist when parts of a whole are considered as a whole *by an observer*. Why is this significant? The reader is encouraged to take note of the following points:

1. The modelled system is *always* part of another system[2]
2. The boundaries of the system are defined by the observer
3. The parameters in model the are variables on a different time-scale
4. (The observer is a complex system)
5. (The observer can recognise emergent phenomena)

The implications of the latter two points will be returned to, while the former three appear to lead to the conclusion that structure is quite literally an illusion: as the temporal scale

[2]This must applies to 'real-world' phenomena. One possible exception may be a model of the entire universe, but for living systems this case certainly appears to hold.

is changed (increased), all parameters will either eventually change, becoming variables, or cease to exist altogether. To illustrate this consider as an example the evolution of the system described by Equation 1, where initially $y = 10$ (see Figure 1 below). When observing from one time-frame the system appears stable to positive and negative perturbation, approaching a fixed-point attractor. In this case p can be thought of as a parameter initialised to zero, since this will make negligible difference to the observed behaviour of the system at this time-scale. However, on a larger time-scale, it is apparent that the attractor is actually in motion, and thus the parameter p is in fact a variable. This is not as counter-intuitive as one might think: try imagining a physical structure that will never change.

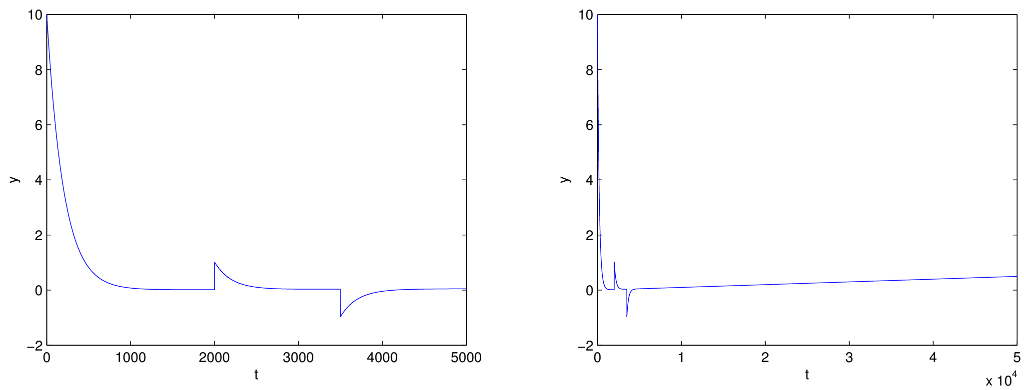

Figure 1: Change in variable y over time, t, shown from two different temporal 'perspectives'. The system is perturbed positively and negatively at time 2000 and 3500 respectively. The shifting of the stable point is not apparent in the smaller time-frame (left).

A human being can therefore only *approximately* model a subset of a system within a specific spatio-temporal frame. It also implies that objectively 'form and function' are identical, as they are both structural change on different time-scales. It is only sensible to then assume that it is the dynamics of the structures that constitute us as observers that allow a subjective definition of change and invariance, together with a sense of time. Furthermore, our bodies (embodiment) must impose particular structure on sensory stimuli (perturbations), providing us with a stable 'point of view' from which to abstract topological features. Considering this, since humans share similarly structured bodies and brains governed by the same developmental principles, it may make more sense to say this in fact inter-subjective, rather than subjective, as then must be our reality.

When development, learning, adaptation and coordination are analysed from this perspective, they are variations of the same thing: structure that resists perturbation. This casts development and *innateness* in a new light, as a continuous measure of resistance to perturbation, closely relating to probability. Probability can then be thought of as the persisting existence of a structure at a particular point in time, despite perturbation to the system at a previous time. For example, the 'probability' of a human developing two arms and two legs is extremely high during the process of embryogenesis, but only since the development of such structures are highly resistant to many kinds of environmental perturbation. In contrast, other structures, such as the central nervous system are more sensitive to perturbation. In fact, certain of such structures *require* or exploit perturbation: it is well known that early sensory deprivation (lack of perturbation) will lead to maldevelopment in the human brain. Structural neural plasticity is a further example, where perturbations (sensory stimuli) are exploited to change the structure of the brain in order to facilitate some particular task. From here it is a trivial step to imagine how resisting and exploiting perturbations in the right way may allow a structure to form a kind of 'history', or memory.

The perspective can be further extended by translating many different high-level cognitive phenomena into the same language. It is hoped that in doing so eventually conclusions could be drawn by recognising commonalities and inferring salient dynamics. Why is it that we seem to be able to recognise emergent phenomena beyond the niche in which we evolved? The laws of physics appear to give rise to complex laws (or vice versa), such as self-organisation, which apply to multiple spatio-temporal scales. As a result different phenomena appear to behave in a similar way to each other, emulating one another, or at least allow us to perceive some apparent emulation or similarity, e.g., spirals, geometric structures, etc. Perhaps it is this that we are exploiting. When considering the problem of instantiating the dynamic responsible for open-ended learning, the language can not only be used to help us design more realistic systems (as detailed in the following section), but it can also be used in a completely different way, leading to different questions, e.g., What does language tell us about the patterns we can categorise? What does the presence of "emergent phenomena" in our vocabulary signify?

4 Intelligence Without Algorithms

Having set up a more abstract perspective from which to approach the problem, the question of open-ended skill acquisition and integration changes (approximately) from this:

> "What architecture should I use so my system can (learn to) do N, P, Q ... in situations A, B, C? How can I allow the system to (learn to) represent knowledge or a skill? How can these skills then be integrated?"

to this:

> "What kind of topological *structure* and laws of *change* should constitute my system, such that when it is coupled to the environment it displays dynamic X, thus (with high probability eventually) achieving Y?".

Where N, P, Q, etc., are high-level descriptions of cognitive phenomena and A, B, C, etc., are high-level descriptions of situations or contexts. In the latter case X is a developmental dynamic, Y is an appropriately low-level description of the desired behaviour(s) translated into the language advocated by the perspective. The "environment"" *includes* some kind of morphology, e.g., a humanoid robot, and the "with high probability eventually" refers to the continuous innateness as discussed in the previous section.

Notice that the first question can be translated into the second, and that part of the problem now becomes defining 'open-ended skill acquisition' in the new perspective. In fact, in this lower-level language the definition will inevitably be more complicated, to the extent where it may assist in recreating the phenomena it describes. Perhaps this is because the language is closer to the 'essence' of the phenomena, or the mathematical model that describes it. By modelling something, we are typically able to predict it. Is this a cooincidence? Lower-level language such as "moving your hand left and right" in comparison to a higher level, "waving", appear to have a closer association with the degrees of freedom of the phenomena they are 'modelling'.

Specifying the problem in this way also avoids task-specification and programming complicated algorithms yielding behaviour restricted to potentially unanticipated domains. There is no symbol grounding problem (Harnad 1990) because there are no symbols to ground and adaptivity may be provided 'for free' by ultrastable dynamics (Iizuka & Paolo 2008, Ashby 1960). The system is autonomous, and there is no integration problem, 'combinatorial explosions' or intractable search-spaces because there is nothing to integrate or enumerate combinatorially. The initial structure is simple and relies on the complexity of the environment in order to 'complexify', changing its structure on different time-scales in order to learn and develop. This *is* the nature of living systems, and it may even be the case that *only* by simulating it more directly will we be able to emulate its fundamental principles; we must capture the salient degrees of freedom.

However, one could argue that re-frameing a problem does not change it in any way - *how exactly* shall we choose an appropriate initial structure, law of change and morphology? What might this involve?

4.1 Doing Something Rather Than Nothing

While there may be no complicated architectures to design or components to construct, we now face with a different problem: namely, that we cannot see into the future. Rapid prototyping is difficult or impossible since the evolution of the system during its ontogenic development, including whether or not it functions as desired, relies as much on the changing structure of the environment in *real-time* as the structure of the system and the rules that determine how it changes. Another way of looking at the problem is that we would like to maximise the *probability* or continuous innateness of the system developing in the 'right way'. Fortunately related work is already taking place in developmental biology, developmental psychology and particularly developmental (epigenetic) robotics. The approach advocated here could in fact be seen as a subset of developmental robotics.

Developmental stages or *bootstrapping* is one possible way of ensuring a system develops as required. In the context of new perspective, developmental stages and bootstrapping can be understood as co-dependent structures towards to higher (more likely, more resistant to perturbation) end of the continuous innateness scale, such that each structure increases the probability of a subsequent one emerging. The ontogenic time may then be divided in several ways, since the order of the stages can be flexible. In human infants developmental stages do not always occur in the same order, but they do *reliably occur* (Thelen & Smith 1996). The designer of the system must understand the principles of this process in order to create the correct initial conditions, looking at the system as both a continuously changing structure, but progressing through a series of qualitatively different observer-dependent stages. Computational work into the dynamics underlying developmental stages for implementation in artificial systems is already being carried out (Lee et al. 2007) based on studies of sensorimotor learning in infants.

High and low-level studies of phenomena such as motivation, self-exploration and curiosity will also likely be necessary to provide the system with some kind of drive or *intrinsic motivation* so that it does "something rather than nothing" (Oudeyer et al. 2007, Der et al. 2005), driving it through develoopmental stages and eventually towards learning by imitation or 'intelligent others' (Smith & Breazeal 2007). Although still rather algorithmic, a notable project by Oudeyer et al. (2005) referred to as *The Playground Experiment* featured a Sony AIBO robot learning 'sensorimotor affordances' without any assistance or pre-programming: the intrinsic dynamics take the form of a mechanism which forces the robot into situations that are "neither too predictable nor too unpredictable" in order to maximise its learning progress. Behaviour of increasing complexity emerges and the robot progresses through several 'phases', from motor-babbling to tracking objects, to coordinated motor primitives towards particular objects. This domain of work within robotics is also sometimes referred to as artificial curiosity (see Stojanov & Kulakov (2006) for a review).

Prince et al. (2005) also address this issue and report that a number of authors have suggested the need for a 'self', and state that self-other discrimination may provide a means to 'bootstrap the skills of a developing robot'. Smith & Breazeal (2007) suggest discuss 'coupling to intelligent others' as a key principle of development, in addition to two other principles that may potentially lead to more 'human-like' intelligence. Artificial evolution may also provide another approach to understanding the principles of development. For example, recent evolutionary robotics research using modulated recurrent neural networks as control circuits for embodied agents (Iizuka & Paolo 2008). Although the author does not specifically describe this work as developmental research, certain parameters in the evolved network rely on environmental perturbations to self-stabilise, and as a result are robust to morphological disruption.

5 Conclusion

In this paper an attempt was made to create a more abstract perspective acknowledging the observer-dependence of different cognitive phenomena. The advantages of using a more abstract, lower-level terminology were emphasised as two-fold: to produce models capturing the salient degrees of freedom of living systems over an ontogenic scale in order to create general artificial intelligence systems, and to make deeper connections between cognitive phenomena usually considered qualitatively different to one another, thus leading to new questions and thus new answers.

Cognitive architectures with fixed processing structures were shown to be fundamentally different to the living systems they attempt to model, and as a result it was concluded that they may fail to capture the 'essence', or key dynamics necessary for what we observe as the open-ended acquisition and integration of new skills. Truly open-ended adaptive general intelligence may only be achievable using self-organising structures 'programmed by nature'.

References

Ashby, R. (1960), *Design for a Brain*, Chapman & Hall, London.

Brooks, R. A. (1991*a*), Intelligence without reason, *in* J. Myopoulos & R. Reiter, eds, 'Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)', Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, Sydney, Australia, pp. 569–595.

Brooks, R. A. (1991*b*), 'Intelligence without representation', *Artificial Intelligence* **47**(1-3), 139–159.

Brooks, R. A. (1996), 'From earwigs to humans', *Robotics and Autonomous Systems* **20**, 291–304.

Brooks, R. A., Breazeal (Ferrell), C., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B. & Williamson, M. M. (1998), Alternative essences of intelligence, *in* 'AAAI '98/IAAI '98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence', American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 961–968.

Der, R., Hesse, F. & Martius, G. (2005), 'Rocking stamper and jumping snake from a dynamical system approach to artificial life', *J. Adaptive Behavior* **14**, 116.

Haken, H. P. J. (1988), 'Synergetics', *Circuits and Devices Magazine, IEEE* **4**(6), 3–7.

Harnad, S. (1990), 'The symbol grounding problem', *Physica D* **42**, 335–346.

Iizuka, H. & Paolo, E. A. (2008), Extended homeostatic adaptation: Improving the link between internal and behavioural stability, *in* 'SAB '08: Proceedings of the 10th international conference on Simulation of Adaptive Behavior', Springer-Verlag, Berlin, Heidelberg, pp. 1–11.

Kirsh, D. (1991), 'Today the earwig, tomorrow man?', *Artif. Intell.* **47**(1-3), 161–184.

Lee, M. H., Meng, Q. & Chao, F. (2007), 'Staged competence learning in developmental robotics', *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems* **15**(3), 241–255.

Lungarella, M. (2007), 'Developmental robotics', *Scholarpedia* **2**(8), 3104.

Lungarella, M., Metta, G., Pfeifer, R. & Sandini, G. (2003), 'Developmental robotics: a survey', *Connection Science* **15**(4), 151–190.

Metzinger, T. (2007), 'Self models', *Scholarpedia* **2**(10), 4174.

Newell, A. (1994), *Unified theories of cognition*, Harvard University Press, Cambridge, MA, USA.

Oudeyer, P.-Y., Kaplan, F. & Hafner, V. (2007), 'Intrinsic motivation systems for autonomous mental development'.

Oudeyer, P. Y., Kaplan, F., Hafner, V. V. & Whyte, A. (2005), 'The Playground Experiment: Task-Independent Development of a Curious Robot', *Proceedings of the AAAI Spring Symposium on Developmental Robotics* pp. 42–47.

Prince, C., Helder, N. & Hollich, G. (2005), 'Ongoing emergence: A core concept in epigenetic robotics'.

Serre, T., Oliva, A. & Poggio, T. (2007), 'A feedforward architecture accounts for rapid categorization', *PNAS* **104**(15), 6424–6429.

Sloman, A. (1999), What sort of architecture is required for a human-like agent?, *in* 'Foundations of Rational Agency', Kluwer Academic, pp. 35–52.

Sloman, A. (2008), Putting the Pieces Together Again, *in* R. Sun, ed., 'Cambridge Handbook on Computational Psychology', Cambridge University Press, New York, chapter 26, pp. 684–709. http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#710.

Smith, L. & Breazeal, C. (2007), 'The dynamic lift of developmental process', *Developmental Science* **10**, 61–68.

Stojanov, G. & Kulakov, A. (2006), 'On curiosity in intelligent robotic systems'.

Thelen, E. & Smith, L. B. (1996), *A Dynamic Systems Approach to the Development of Cognition and Action (Cognitive Psychology)*, The MIT Press.

Vernon, D., Metta, G. & Sandini, G. (2007), 'A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents', *Evolutionary Computation, IEEE Transactions on* **11**(2), 151–180.

Wang, P., Goertzel, B. & Franklin, S., eds (2008), *Artificial General Intelligence 2008, Proceedings of the First AGI Conference, AGI 2008, March 1-3, 2008, University of Memphis, Memphis, TN, USA*, Vol. 171 of *Frontiers in Artificial Intelligence and Applications*, IOS Press.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M. & Thelen, E. (2001), 'Artificial intelligence. autonomous mental development by robots and animals.', *Science* **291**(5504), 599–600.