# Cognitive Vision – The Development of a Discipline

David Vernon

**The following views regarding Cognitive Vision are excerpts from David Vernon´s article „Cognitive Vision – The Development of a Discipline", available for download from http://www.ecvision.org/about_ecvision/Definition_of_Cognitive_Vision.htm**

## The Balance between Phylogeny and Ontogeny: Hard-Wired Functionality vs. Learned Capabilities

In cognitive systems, one often distinguishes between phylogeny and ontogeny. Phylogeny refers to the initial configuration of the system and its evolution from generation to generation. On the other hand, ontogeny refers to the learning and development of a given system during its lifetime. The issue that arises in this context is the requirements for the minimal architecture for a cognitive vision system. There are two perspectives on this, depending on whether one takes a cognitivist stance or an emergent stance.

In the cognitivist stance, the issue comes down to the balance between required 'pre-knowledge' and acquirable knowledge. Or, put another way, how much does one need to know and to be able to do in order to be capable of learning new things, such as concepts or actions? That is, we need a clear cut set of conditions under which certain learning can take place.

In the emergent stance, there is a trade-off between phylogenic configuration and ontogenic development. Phylogeny determines the visuo-motor capability that a system is configured with at the outset and which facilitates the system's innate behaviours. Ontogenic development gives rise to the cognitive capabilities that we seek. Since we don't have the luxury of having evolutionary timescales to allow phylogenic emergence of a cognitive system - we can't wait around to evolve a cognitive system from nothing - we must somehow identify a minimal phylogenic state of the system. In practice, this means that we must identify and effect visuomotor capabilities for the minimal reflex behaviours that ontogenic development will subsequently build on to achieve cognitive behaviour. Put simply: we need to decide what visual processing capabilities are needed for a minimal emergent cognitive vision system.

## The Necessity of Embodied Cognition

The question as to whether cognitive vision systems have to be physically embodied or not is one of the most contentious and divisive issues in the field. The divisiveness arises from the different stances taken by the different paradigms.

From the perspective of the cognitivist paradigm, there is actually no case for embodiment, at least none for it as a mandatory requirement of cognition. Cognitivist systems don't necessarily have to be embodied. The very essence of the cognitivist approach is that cognition comprises computational operations defined over symbolic representations and these computational operations are not tied to any given instantiation. They are abstract in principle. It is for this reason that it has been noted that cognitivism exhibits a form of mind-body dualism [12,27]. Symbolic knowledge, framed in the concepts of the designer, can be programmed in directly and doesn't have to be developed by the system itself through exploration of the environment. Some cognitivist systems do exploit learning to augment or even supplant the a priori designed-in knowledge and thereby achieve a greater degree of adaptiveness, reconfigurability, and robustness. Embodiment may therefore offer an additional degree of freedom to facilitate this learning, but it is by no means necessary. The clear advantage of this position is that a successful cognitivist model of cognition could be instantiated in any context and, theoretically at least, be ported to any application domain.

The perspective from emergent systems is diametrically opposed to the cognitivist position. Emergent systems, by definition, must be embodied and embedded in their environment in a situated historical developmental context [12]. To see why embodiment is a necessary condition of emergent cognition, consider what cognition means in the emergent paradigm. It is the process whereby an autonomous system becomes viable and effective in its environment. In this, there are two complementary things going on: one is the self-organization of the system as distinct entity1, and the second is the coupling of that entity with its environment. 'Perception, action, and cognition form a single process' [27] of self-organization in the specific context of environmental perturbations of the system. This gives rise to the co-development of the cognitive system and its environment and thereby to the ontogenic development of the system itself over its lifetime. This development is identically the cognitive process of establishing the space of mutually-consistent couplings. Put simply, the system's actions define its perceptions but subject to the strong constraints of continued dynamic self-organization. The space of perceptual possibilities is predicated not on an 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



**Korrekturabzug:** Künstliche Intelligenz, Heft 2/05,
**per fax**: +49 421 1630 1952
technische Hinweise unter **www.kuenstliche-intelligenz.de/ /Thema/HinweisefuerAutorInnenundGast-Herausgebe.htm**

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 ontogenic development of the system. Through this ontogenic development, the cognitive system develops its own epistemology, i.e.its own system-specific 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. Thus, we can see that, from this perspective, cognition is inseparable from 'bodily action' [27]: without physical embodied exploration, a cognitive system has no basis for development.

## A Brief Overview of Paradigms in Cognition and Cognitive Vision

There are several distinct approaches to the understanding and synthesis of cognitive systems. These include 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.

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, 11, 12, 5]. Traditionally, this has been the dominant theme in cognitive science [8] but there are indications that the discipline is migrating away from its stronger interpretations [6].

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 an objective external world [13]. Reasoning itself is symbolic: a procedural process whereby explicit representations of an objective world are manipulated and possibly translated into language.

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.

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.

Co-determination is one of the key differences between the emergent paradigm and the cognitivist paradigm. For emergent systems, perception 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, cognition and perception is functionally-dependent on the richness of the action interface [17].

Dynamical systems theory is one of the most promising approaches to the realization of emergent cognitive systems. Advocates of the dynamical systems approach to cognition (e.g. [8, 12, 18]) argue that motoric and perceptual systems, as well as perception-action coordination, are dynamical systems, that self-organize into meta-stable patterns of behaviour.

Proponents of dynamical systems point to the fact that they directly provide 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.

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 [18, 5], enaction [19, 20, 21, 16, 22, 5, 13] asserts that 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 symbolic representations. Instead there is an enactive interpretation: a real-time context-based choosing of relevance. The advantage is that it focusses on the dynamics by which robust interpretation and adaptability arise.

Recently, effort has gone into developing approaches which combine aspects of the emergent systems and cognitivist systems [17, 4, 23]. These hybrid approaches have their roots in strong criticism of the use of explicit programmer-based knowledge in the creation of artificially-intelligent systems [24] and in the development of active 'animate' perceptual systems [25] 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' [4]. Hybrid systems are in many ways consistent with emergent systems while still exploiting programmer-centred (but not programmer-populated) representations (for example, see [26]).

It is important to be aware that the different paradigms of cognitive vision are not equally mature and it isn't clear which paradigm will ultimately be successful. The arguments in favour of dynamical systems and enactive systems are compelling but, though they offer great promise, the current capabilities of cognitivist systems are actually more advanced. However, they are also quite brittle and have achieved little in the cognitive capabilities associated with generalization. Enactive and dynamical systems should in theory be much less brittle because they emerge through mutual specification and co-development with the environment, but their cognitive capabilities are actually very limited at present. The extent to which this will change and the speed with which this change will occur is uncertain. Hybrid approaches seem to offer the best of both worlds but it is unclear how well one can combine what are ultimately highly antagonistic underlying philosophies.

## Advances in Related Disciplines

The discipline of cognitive vision is developing in the context of other disciplines which are themselves evolving and changing. New insights will come not only from the cognitive vision community but also from the broader multi-disciplinary community. We note here just two examples of how advances in related disciplines will have an impact on our own efforts to create the solid foundations of visually-enabled cognitive systems. The disciplines involved are cognitive science, neuroscience, and epigenetic robotics.

In the last 10 years or so, an ever growing number of cognitive scientists [28, 29] have begun to appreciate the possibility of instantiating cognitive models in robotic systems. The space of research spanned is quite wide, starting from the locomotion and organizational behaviors of insects and early vertebrates [30, 31] through models of high order cognitive skills in humans such as social behaviors [32], imitation [33, 34, 35], communication, and language [34, 36, 37, 38, 39]. More recently a new strain of research explicitly included developmental aspects and the modeling of development [40, 41] and epigenetic robotics [40]. Examples are the work of Metta and Sandini [42, 43, 44, 45], of the group of Pfeifer [46, 47, 48], and of Dautenhahn et al. [49, 50].

Imitation is one of the key stages in the development of more advanced cognitive capabilities. While the study of infants and adults ability to imitate has remained foremost a field of the psychological literature, recently, it has found a ground in the neurological literature with the discovery of the mirror neuron system in monkeys [51]. The mirror neuron system is formed by pre-motor neurons discharging both when the animal acts and when it sees similar actions performed by other individuals. A system, similar to that found in monkeys, has been indirectly shown to exist also in humans by transcranial magnetic stimulation studies of the motor cortex during action observation [52]. Further investigations have shown that the mirror system can be activated not only by visually perceived actions but also by listening to action-related sounds [53] and, in humans, by speech listening [54]. In addition to these electrophysiological data, in humans, a number of brain imaging studies point all to a network of brain areas responsible for the visuo-motor transformation mechanism underlying action recognition [55, 56]. It is plausible that the motor resonant system formed by mirror neurons is involved in someone else's action understanding and, at least in humans, imitation.

## References


[1] O. D. Faugeras. *3-D Computer Vision.* MIT Press, 1993.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[3] T. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and B. Neumann, editors, *Proceeding of the 5th European Conference on Computer Vision*, volume 2, pages 484–498. Springer-Verlag, 1998.

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

[5] 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*, volume 130 of Boston Studies in the Philosophy of Science, pages 235–263. Kluwer Academic Publishers, 1992.

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

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

[8] J. A. S. Kelso. *Dynamic Patterns – The Self-Organization of Brain and Behaviour*. M.I.T. Press, 3rd edition, 1975.

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

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

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

[12] 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.

[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*, 23:803–817, 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] 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*, Cambridge, Massachusetts, 1995. Bradford Books, MIT Press.




**Korrekturabzug:** Künstliche Intelligenz, Heft 2/05,
**per fax**: +49 421 1630 1952

technische Hinweise unter **www.kuenstliche-intelligenz.de/ /Thema/HinweisefuerAutorInnenundGast-Herausgebe.htm**

[19] H. Maturana. *Biology of Cognition*. 1970.
[20] H. Maturana. *The Organization of the Living: a Theory of the Living Organization*. 1975.
[21] H. R. Maturana and F. J. Varela. *Autopoiesis and Cognition*, volume 42 of Boston Studies on the Philosophy of Science. D. Reidel Publishing Company, 1980.
[22] F. Varela. *Principles of Biological Autonomy*. 1979.
[23] G. H. Granlund. *Cognitive vision – background and research issues*. Linkoping University, 2002.
[24] H. L. Dreyfus. From micro-worlds to knowledge representation. In J. Haugland, editor, *Mind Design: Philosophy, Psychology, Artificial Intelligence*, 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.
[25] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
[26] 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, *Proceedings of the 8th European Conference on Computer Vision*, ECCV 2004, volume I of LNCS, pages 28–39. Springer, 2004.
[27] E. Thelen. Time-scale dynamics and the development of embodied cognition. In R. F. Port and T. van Gelder, editors, *Mind as Motion – Explorations in the Dynamics of Cognition*, pages 69–100, Cambridge, Massachusetts, 1995. Bradford Books, MIT Press.
[28] G. M. Edelman. *Neural Darwinism: The Theory of Neuronal Group Selection*. Oxford University Press, Oxford, 1988.
[29] M. I. (Ed.) Posner. *Foundations of Cognitive Science*. The MIT Press, Cambridge, MA, fifth edition, 1996.
[30] A. J. Ijspeert. Synthetic approaches to neurobiology: review and case study in the control of anguiliform locomotion. In *Fifth European Conference on Artificial Life (ECAL99)*, 1999.
[31] R. Moller, M. Marinus, and D. Lambrinos. A neural model of landmark navigation in insects. In *Computational Neuroscience Meeting CNS'98*, 1998.
[32] C. Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Unpublished Doctoral Dissertation. MIT, Cambridge, MA., 2000.
[33] M. Arbib, A. Billard, M. Iacoboni, and E. Oztop. Mirror neurons, imitation and (synthetic) brain imaging. *Neural Networks*, 13(8/9):953–973, 2000.
[34] A. Billard. Imitation. In M.A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, 2002.
[35] S. Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3(6):233–242, 1999.
[36] A. Billard and K. Dautenhahn. Grounding communication in autonomous robots: an experimental study. *Robotics and Autonomous Systems*, 24(1/2 "Scientific methods in mobile robotics"):71–79, 1998.
[37] A. Billard and K. Dautenhahn. Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive Behavior*, 7(3/4):415–438, 2000.
[38] L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In P. Husbands and I. Harvey, editors, *Fourth European Conference on Artificial Life,* ECAL97, pages 473–484. MIT Press, 1997.
[39] P. Varshavskaya. Behavior-based early language development on a humanoid robot. In *Epigenetic Robotics*, pages 149–158, Edinburgh, UK, 2002. Lund University.
[40] J. Zlatev and C. Balkenius. Introduction: Why "epigenetic robotics"? In *First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, volume 85, pages 1–4. Lund University Cognitive Studies, 2001.
[41] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. Developmental robotics: A survey. *Connection Science*, Forthcoming, 2003.
[42] G. Metta. *Babybot: a Study on Sensori-motor Development*. Ph.d. thesis, University of Genova, 2000.
[43] G. Metta and P. Fitzpatrick. Early integration of vision and manipulation. *Adaptive Behavior*, 11(2):109–128, 2003.
[44] G. Metta, G. Sandini, and J. Konczak. A developmental approach to visually-guided reaching in artificial systems. *Neural Networks*, 12(10):1413–1427, 1999.
[45] G. Metta, G. Sandini, L. Natale, and F. Panerai. Development and robotics. In *IEEE-RAS International Conference on Humanoid Robots*, Tokyo, Japan, 2001. Humanoids2001 organizers.
[46] M. Lungarella and L. Berthouze. Adaptivity through physical immaturity. In *2nd International Workshop on Epigenetic Robotics (EPIROB'02)*, pages 79–86, Edinburgh, Scotland, 2002.
[47] R. Pfeifer and C. Scheier. Sensory-motor coordination: The metaphor and beyond. *Robotics and Autonomous Systems*, 20:157–178, 1997.
[48] R. Pfeifer and C. Scheier. Representation in natural and artificial agents: an embodied cognitive science perspective. In *Natural Organisms, Artificial Organisms, and Their* Brains, pages 480–503, Bielefeld, Germany, 1998. Verlag.
[49] K. Dautenhahn and S. Coles. Narrative intelligence from the bottom up: A computational framework for the study of story-telling in autonomous agents. *JASSS, The Journal of Artificial Societies and Social Simulation*, 2000.
[50] K. Dautenhahn, B. Ogden, and T. Quick. From embodied to socially embedded agents – implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428, 2002.
[51] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131–141, 1996.
[52] L. Fadiga, L. Fogassi, G. Pavesi, and G. Rizzolatti. Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73(6):2608–2611, 1995.
[53] E. Kohler, C. Keysers, M. A. Umilta, L. Fogassi, V. Gallese, and G. Rizzolatti. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297:846–848, 2002.
[54] L. Fadiga, L. Craighero, G. Buccino, and G. Rizzolatti. Speech listening specifically modulates the excitability of tongue muscles: a tms study. *European Journal of Neuroscience*, 15:399, 2002.
[55] J. Decety, T. Chaminade, J. Grezes, and A.N. Meltzo_. A pet exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage*, 15:265–272, 2002.
[56] M. Iacoboni, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, and G. Rizzolatti. Cortical mechanisms of human imitation. *Science*, 286:2526–2528, 1999.

**Kontakt?**
**Foto?**
**Vita?**