ROBotic Open-architecture Technology for
Cognition, Understanding, and Behavior

Project No. 004370

RobotCub

Development of a Cognitive Humanoid Cub

Instrument: Integrated Project
Thematic Priority: IST – Cognitive Systems

D2.1 A Roadmap for the Development of Cognitive Capabilities in Humanoid Robots

Due Date: **31/01/2010**
Submission date: **30/12/2009**

Start date of project: **01/09/2004** Duration: **65 months**

Organisation name of lead contractor for this deliverable: **University of Genoa**

Responsible Person: **David Vernon**

Revision: **6.5**

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary

RobotCub is a research project dedicated to the investigation of cognitive systems through the ontogenic development[1] of a humanoid robot. That is, it is a programme of enquiry into emergent embodied cognitive systems whereby a humanoid robot, equipped with a rich set of innate action and perception capabilities, can develop over time an increasing range of cognitive abilities by recruiting ever more complex actions and thereby achieving an increasing degree of prospection (and, hence, adaptability and robustness) in dealing with the world around it.

Cognitive development involves several stages, from coordination of eye-gaze, head attitude, and hand placement when reaching, through to more complex — and revealing — exploratory use of action. This is typically achieved by dexterous manipulation of the environment to learn the affordances of objects in the context of one's own developing capabilities. Our ultimate goal is to create a humanoid robot — the iCub — that can communicate through gestures simple expressions of its understanding of its environment, an understanding that is achieved through rich manipulation-based exploration, imitation, and social interaction.

Deliverable D2.1 encapsulates several contributions to the eventual creation of a model of cognition and an associated architecture which will facilitate the development of a spectrum of cognitive capabilities in the iCub humanoid robot. It comprises five parts.

Part I presents a conceptual framework that forms the foundation of the RobotCub project, identifying the broad stance taken in the project to cognitive systems — emergent embodied systems that develop cognitive skills as a result of their action in the world — and drawing out explicitly the strong consequences of adopting this stance.

Part II surveys what is known about cognition in natural systems, particularly from the developmental standpoint, with the goal of identifying the most appropriate system phylogeny and ontogeny.

Part III explores neurophysiological and psychological models of some of these capabilities, noting where appropriate architectural considerations such as sub-system interdependencies that might shed light on the overall system organization.

Part IV then sets out to provide a synopsis of the current models that the RobotCub partners are working with. It places them in a two-dimensional space of ontogeny, spanned by actions and prospective capabilities, that is traversed by a cognitive system as it develops from its initial phylogenically-endowed state towards greater cognitive ability, such as imitation and communication (and, by extension, deliberation and reasoning).

Part V presents a roadmap that uses the phylogeny and ontogeny of natural systems to define the innate skills with which the humanoid robot must be equipped so that it is capable of ontogenic development, to define the ontogenic process itself, and to show exactly how the humanoid robot should traverse the two-dimensional space of ontogeny. Part V concludes by setting out an agenda for subsequent research and addresses the creation of an architecture for cognition: a framework for operational integration of discrete capabilities and the challenge of theoretical unification of distinct models.

This deliverable will be produced incrementally over an extended period; of the five parts comprising the document, only Parts I, II, and V are substantially complete at this time, with significant progress having been made in Part III.

[1] We qualify *ontogenic development* to distinguish it from the technological development of the mechatronic humanoid robot itself: once the humanoid robot has been designed and constructed, it will be used then to study cognition by letting the robot itself develop cognitive skills through its interaction with people and the world around it.

Principal Contributors

Giorgio Metta, University of Genoa

Laila Craighero, University of Ferrara

Luciano Fadiga, University of Ferrara

Auke Ijspeert, EPFL

Giorgio Metta, University of Genoa

Kerstin Rosander, University of Uppsala

Giulio Sandini, University of Genoa

David Vernon, University of Genoa

Claes von Hofsten, University of Uppsala

Revision History

Version 1.0 (DV 13-09-2005)
Original draft.

Version 1.1 (DV 20-09-2005)
Reworked Executive Summary and elaborated the structure (but not yet the content) of Parts III, IV, and V.

Version 1.2 (DV 01-10-2005)
Most references in Part II are now cited correctly. Content added to Part III. A note added to the Executive Summary to make the incremental nature of the deliverable explicit.

Version 1.3 (DV 08-10-2005)
Introduction to Executive Summary extended. List of principal contributors added. Miscellaneous clarifications made in Parts I and II. Section on prospection in Part IV removed (prospection is addressed implicitly in all the other sections in this Part). Various grammatical errors corrected.

Version 1.4 (DV 12-10-2005)
Introduction to Executive Summary extended further. Summary of Part III added.

Version 2.0 (DV 19-04-2006)
Part II restructured and new sections added on core knowledge (Section 6.1), visual development (Section 7.3), space perception (Section 7.3.1), and the summary of innate core abilities of the neonate (Section 9.2). New material added in Part V, specifically on the baseline phylogenetic configuration of the iCub (Section 15.6.1), a proposed cognitive architecture (Section 15), and the experimental scenarios for the ontogenetic development of the iCub (Section 16),

Version 2.1 (DV 7-06-2006)
Part V updated: new section contrasting the different paradigms of cognition, added new examples to the survey of cognitive architectures, added a comparative analysis, and expanded the section on implications for development of cognition.

Version 3.0 (DV 29-07-2006)
Part V updated: the possibility of interconnection between phylogenetic skills, the identification of three new skills, and priority for implementation (Section 15.6.1), clarification of the biological inspiration of the components of the modulation circuit in the cognitive architecture (Section 15.6.2 and Figure 10), the relevance of action selection, recognition, inference, and communication in the prospection circuit (Section 15.6.3), new scenario on learning to crawl (Section 16).

Version 4.0 (DV 07-04-2007)
Part V updated: A rationalized set of phylogentic capabilities has been identified in Section 15.6.1 and a set of empirical investigations has been added in Section 17.

Version 5.0 (DV 04-10-2007)
Part V updated: The further development of the cognitive architecture is described in Section 15.6.5 and in particular initial steps towards its realization are described in Section 15.6.6.

Version 6.0 (DV 30-01-2008)
Converted from LaTeX to Microsoft Word to facilitate independent revision by the principal contributors. It was necessary to allow several idiosyncrasies to facilitate this conversion. All lines are terminated by an end-of-line character to preserve the original formatting, resulting in a ragged right margin. Text introduced in subsequent revisions need not follow this convention. The document version number and

date must be updated independently on both the front cover and the footer at the end of the page. The original section numbers, table numbers, and figure numbers have been retained but as literal values; they are not updated automatically. New sections, tables, and figures must be numbered appropriately. References in the body of the text to section, figure, and table numbers are also retained, but again as literal values. Citations have been converted from numbers, e.g. [1], to alpha-style, e.g. [Zie03] for single author, [TS94] for two authors, [TKS95] for three authors, and [SME+04] for four or more authors. These are listed in alphabetic order in the references at the end of the document. New references should be inserted at the appropriate point and cited using the convention noted above.

Version 6.1 (DV 2-06-2008)
Part V, Section 17: Empirical Investigations updated. Existing material has been organized into five sub-sections dealing with *Looking*, *Reaching*, *Reach and Grasp*, *Reach and Posture*, and *Postural Control* in Action. Some minor amendments have been made to these sub-sections, most notably Section 17.3 Reach and Grasp, points 3 and 6. New material has been introduces in two additional sub-sections dealing with *Object Containment* and *Pointing and Gesturing*. Finally, a new sub-section entitled *A Comprehensive Experiment* has been added to demonstrate the integration of all work-packages.

Version 6.2 (DV 18-10-2008)
Added material to Part II, Section 6, *Core Abilities*: in particular to to Section 6.1, *Objects*, and 6.4, *People*. New section added to Part II: Section 7.3.2 *Object Perception*. Material added to Part II, Section 7.4.1 *Development of Posture and Locomotion* and Section 7.4.3 *Development of Reaching and Manipulation* (four new subsections on *Reaching*, *Grasping, Bimanual Coordination,*and *Manipulation*). Material added to Section 7.4.4, *Development of Social Abilities*. These amendments are shown in blue typeface to allow them to be easily identified.

Version 6.3 (DV 16-02-2009)
Incomplete references updated.

Version 6.4 (DV 10-05-2009)
Updated Section 17, *Empirical Investigations* by adding a sub-titles to individual aspects of the first three experiments: *Looking, Reaching,* and *Reach and Grasp* (Sections 17.1 – 17.3). These sub-titles attempt to convey the chief characteristics of the experiments.

Version 6.5 (DV 29-12-2009)
Updated Section 15.6, *The iCub Cognitive Architecture*. Extended Section 15.6.6, *Realization of an Essential Phylogeny*. Added Section 15.6.7, *Realization of the iCub Cognitive Architecture*; Section 15.6.8, *Implementation of the Cognitive Architecture*; Section 15.6.9, *The iCub Cognitive Architecture and the Posner Test*; and Section 15.6.10, *Future Work.* Deleted Section 15.7, *Co-Determination and Co-Development Revisited*.

Part I

Scientific Framework

1 iCub — the RobotCub Cognitive Humanoid Robot

The RobotCub project is a research initiative dedicated to the realization of embodied cognitive systems[SMV04b, SMV04a]. It has the twin goals of (1) creating an open humanoid robotic platform for research in embodied cognition — the iCub — and (2) advancing our understanding of cognitive systems by exploiting this platform in the study of the development of cognitive capabilities in humanoid robots. The iCub will have a physical size and form similar to that of a two year-old child and will achieve its cognitive capabilities through development and learning in its environment: by interactive exploration, manipulation, imitation, and gestural communication. The iCub will be a freely-available open system which can be used by scientists in all cognate disciplines from developmental psychology to epigenetic robotics.

As we will see later in this document, one of the tenets of the RobotCub stance on cognition is that manipulation plays a key role in the development of cognitive capability. Consequently, the design is aimed at maximizing the number of degrees of freedom of the upper part of the body (head, torso, arms, and hands). The lower body (legs) will support crawling on arms and legs and sitting on the ground in a stable position with smooth autonomous transition from crawling to sitting. This will allow the robot to explore the environment and to grasp and manipulate objects on the floor. The total number of degrees of freedom is 53 (7 for each arm, 9 for each hand, 6 for the head and 3 for the torso and spine). Each leg will have a further 6 degrees of freedom. The sensory system will include binocular vision and haptic, cutaneous, aural, and vestibular sensors. Functionally, the system will be able to coordinate the movement of the eyes and hands, grasp and manipulate lightweight objects of reasonable size and appearance, crawl using its arms and legs, and sit up. This will allow the system to explore and interact with the environment not only by manipulating objects but also through locomotion.

2 The RobotCub Approach to Cognitive Systems

The RobotCub stance on cognition coincides directly with the emergent systems approach: cognition is the process whereby an autonomous system becomes viable and effective in its environment. It does so through a process of self-organization through which the system is continually re-constituting itself in real-time to maintain its operational identity through moderation of mutual system-environment interaction and co-determination [MV87].

Emergent systems are epitomized by connectionist, dynamical systems, and enactive approaches which view cognition as an emergent property of a network of component elements that comprise a dynamical self-organizing [TS94, Kel95] , self-producing [Mat70, Mat75, Var79, MV80, MV87], self-maintaining system [Mat70, Bic00]. The emergent approach is starkly distinct from the common and prevalent cognitivist standpoint. 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 [CH00b, Ver06, Ver07]. This approach has also been labelled by many as the *information processing* approach to cognition [Mar77, Hau82, Pin84, Kih87, Var92, TS94, Kel95]. In contrast, many emergent approaches assert that 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 [CH00a]. While cognitivism entails a self-contained abstract model that is disembodied in principle, the physical instantiation of the systems plays no part in the model of cognition [Ver07]. In contrast, emergent approaches are intrinsically embodied and, as we will see, the physical instantiation plays a pivotal role in cognition. In the emergent paradigm, there are two complementary issues at stake: one is the coupling of the system with its environment, through perception and action, and the other is the self-organization of the system as a distinct entity. A third issue, embodiment, follows as a consequence of these. In the next section, we will consider each of these issues in turn to see what they imply for the creation of an artificial cognitive system.

3 Requirements for the Realization of Cognitive Systems

3.1 Co-determination: the Requirements of Phylogeny

Co-determination arises from the autonomous nature of a cognitive system. It reflects the fact that an autonomous system[2] defines itself through a process of self-organization and subjugates all other processes to the preservation of that autonomy [Var79]. However, it also reflects the fact that all self-organizing systems have an environment in which they are embedded, from which they make themselves distinct, and which is conceived by the autonomous system in whatever way is supportive of this autonomy-preserving process. In this way, the system and the environment are co-determined: the cognitive agent is determined by its environment by its need to sustain its autonomy in the face of environmental perturbations and at the same time the cognitive process determines what is real or meaningful for the agent, for exactly the same reason. In a sense, co-determination means that the agent constructs its reality (its world) as a result of its operation in that world. Perception provides the requisite sensory data to enable effective action [MV87] but it does so as a consequence of the system's actions, not as a context-free abstraction of information that is descriptive of the world at large [WF86]. Thus, perception is functionally-dependent on the richness of the action interface [Gra99]. Maturana and Varela introduced a diagrammatic way of conveying the self-organized autonomous nature of a co-determined system, perturbing and being perturbed by its environment [MV87]: see figure 1. The arrow circle denotes the autonomy and self-organization of the system, the rippled line the environment, and the bi-directional half-arrows the mutual perturbation.

Co-determination requires then that the system is capable of being autonomous as an entity. That is, it has a self-organizing process that is capable of coherent action and perception: that it possesses the essentials of survival and development. This is exactly what we mean by the phylogenic configuration of a system: the innate capabilities of an autonomous system with which it is equipped at the outset and which form the basis of any subsequent development.

3.2 Co-development: the Requirements of Ontogeny

Co-development, on the other hand, is identically the cognitive *process* of establishing and *enlarging* the possible space of mutually-consistent couplings in which a system can engage (or, perhaps more appropriately, which it can withstand). The space of perceptual possibilities is predicated not on an absolute objective environment, but on the space of possible actions that the system can engage in whilst still maintaining the consistency of the coupling with the environment. These environmental perturbations don't control the system since they are not components of the system (and, by definition, don't play a part in the self-organization) but they do play a part in the ontogenic development of the system. Through this ontogenic development, the cognitive system develops its own epistemology, *i.e.*

[2] Autonomy: the self-maintaining organizational characteristic of living creatures that enables them to use their own capacities to manage their interactions with the world, and with themselves, in order to remain viable [CH00a].

its own system-specific history- and context-dependent knowledge of its world, knowledge that has meaning exactly because it captures the consistency and invariance that emerges from the dynamic self-organization in the face of environmental coupling. Put simply, the system's actions define its perceptions but subject to the strong constraints of continued dynamic self-organization. Again, it comes down to the preservation of autonomy, but this time doing so in an every increasing space of autonomy-preserving couplings.

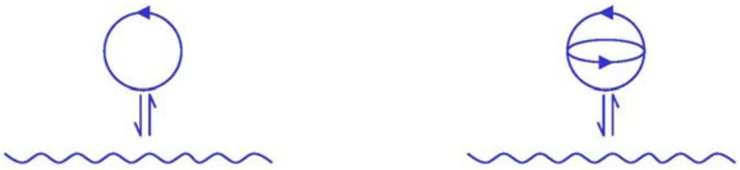


Figure 1: Maturana and Varela's ideograms to denote autopoietic and operationally-closed systems. These systems exhibit co-determination co-development, respectively. The diagram on the left denotes an autopoietic system: the arrow circle denotes the autonomy, self-organization, and self-production of the system, the rippled line the environment, and the bi-directional half-lines the mutual perturbation— structural coupling — between the two. The diagam on the right denotes an operationally-closed autonomous system with a central nervous system. This system is capable of development by means of self-perturbation — self-modification — of its the nervous system, so that it can accommodate a much larger space of effective system action.

This process of development is achieved through self-modification by virtue of the presence of a central nervous system: not only does environment perturb the system (and *vice versa*) but the system also perturbs itself and the central nervous system adapts as a result. Consequently, the system can develop to accommodate a much larger space of effective system action. This is captures in a second ideogram of Maturana and Varela (see figure 1 which adds a second arrow circle to the autopoiesis ideogram to depict the process of self-perturbation and self-modification.

3.3 The Complementarity of Co-determination and Co-development

The system and environment are co-determined (through mutual coupling and contingent self-organization) but some cognitive systems can also adapt through a process of co-development resulting in new co-determined couplings. This complementarity of co-determination and co-development is crucial. We have two distinct but related processes: the co-determination of the system through selforganization in the context of structural coupling (action and perception) and the co-development of the system over time in an ecological and social context as it expands its space of structural couplings (that nonetheless must be consistent with the maintenance of self-organization). Co-development requires additional plasticity of the self-organizational processes. If this is in place, we have both phyogenically-conditioned *co-determination* of the cognitive system and its environment and the potential for ontogenic *co-development* of the system itself over its lifetime.

Co-developement and co-determination together correspond to Thelen's view that perception, action, and cognition form a single process of self-organization *in the specific context of environmental perturbations of the system* [The95]. Thus, we can see that, from this perspective, cognition is inseparable from 'bodily action' [The95]: *without physical embodied exploration, a cognitive system has no basis for development.* Emergent systems, by definition, must be embodied and embedded in their environment in a situated historical developmental context [TS94].

It is important to emphasize that development occurs in a very special way. Action, perception, and cognition are tightly coupled in development: not only does action organize perception and cognition, but perception and cognition are also essential for organizing action. Actions systems do not appear ready-made. Neither are they primarily determined by experience. They result from both the operation of the the central nervous system and the subject's dynamic interactions with the environment. Perception, cognition, and motivations develop at the interface between brain processes and actions. Consequently, advanced cognition (as opposed to homeostatic system stability) is the result of a developmental process through which the system becomes progressively more skilled and acquires the ability to understand events, contexts, and actions, initially dealing with immediate situations and increasingly acquiring a predictive or prospective capability. This dependency on exploration and development is one of the reasons why the iCub requires such a rich space of manipulation and locomotion actions.

We can conclude by noting again that the concept of co-determination is rooted in the Maturana's and Varela's idea of structural coupling of level one autopoietic systems[3] [MV87], is similar to Kelso's circular causality of action and perception each a function of the other as the system manages its mutual interaction with the world [Kel95], and reflect's the organizational principles inherent in Bickhard's self-maintenant systems [Bic00]. The concept of co-development is mirrored in Bickhard's concept of recursive self-maintenance [Bic00] and has its roots in Maturana's and Varela's level two and level three autopoietic systems [MV87].

3.4 Embodiment: the Requirements of Action

If one looks closely at the emergent paradigm, one finds two cornerstones: the operational closure [MV87] (or circular causality [Kel95]) of the system in its self-organization, and the structural coupling of the system with its environment. Operational closure by itself does not imply a need for embodiment: it is an organizational principle and applies to systems of many temporal and spatial scales. Coupling with the environment is a little trickier. The key requirement is that the mutual perturbations implied by the coupling, *i.e.* the mutual system-environment interactions, should be rich enough to drive the ontogenic development but not destructive of the self-organization [MV87]. There is nothing in principle that requires the 'action' to be physical in any strong sense and, therefore, it is possible to develop an embodied cognitive system in any application that offers a suitably rich set of interactions. This is consistent with Ziemke's framework of embodied systems, in which he distinguishes between five types of embodiment (structural coupling, historical embodiment, physical embodiment, organismoid embodiment, and organismic embodiment) [Zie01, Zie03].

There is, however, an important caveat. In a system that only satisfies the minimal requirements of embodiment, there is no guarantee that the resultant cognitive behaviour will be in any way consistent with human models or preconceptions of cognitive behaviour. Of course, this may be quite acceptable, as long as the system performs its task adequately. However, if we want to ensure compatibility with human cognition, then we have to admit the stronger version of embodiment and adopt a domain of discourse that is the same as the one in which we live: one that involves physical movement, forcible manipulation, and exploration, and perhaps even human form [Bro02]. Why? Because when two cognitive systems interact or couple, the shared consensus of meaning — the systems' common epistemology — will only be semantically similar (have similar meaning) if the experiences of the two systems are compatible: phylogenically, ontogenically, and morphologically consistent [MV87]. Consequently, the RobotCub approach to cognition requires that the cognitive systems be embodied in a very specific sense: that it should lie in the organismoid space of embodied cognitive systems and, further still, that it should lie in the humanoid subspace of the organismoid space. Apart from

[3] Autopoiesis is a special type of self-organization: an autopoietic system is a homeostatic system (*i.e.* self-regulating system) but one in which the regulation applies not to some system parameter but to the organization of the system itself [Var79, MV87].

the morphology and phylogeny of the cognitive system, this also has strong implications for the co-development of the cognitive system. Specifically, the ontogeny of the system must follow the development of natural (human) systems. We will deal with this in considerable depth in Parts II and V but it should be noted here that this development follows a general path that begins with actions that are immediate and have minimal prospection, and progresses to much more complex actions that bring forth much more prospective cognitive capabilities. This involves the development of perception/action coordination, beginning with head-eye-hand coordination, progressing through manual and bi-manual manipulation, and extending to more prospective couplings involving inter-agent interaction, imitation, and (gestural) communication.[4] As we will see in Part II, this development occurs in both the innate skills with which phylogeny equips the system and in the acquisition of new skills that are acquired as part of the ontgenic development of the systems. Typically, it is the ontogenic development provides for the greater prospective abilities of cognitive systems.

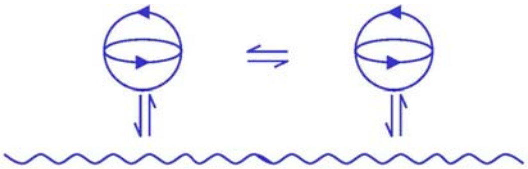


Figure 2: Maturana and Varela's ideogram to denote the co-development engendered by interaction between cognitive systems.

It is important to understand what exactly we mean here by the term 'interaction'. For RobotCub, in adhering to the emergent stance on cognition, interaction is a shared activity in which the actions of each agent influence the actions of the other agents engaged in the same interaction, resulting in a mutually constructed pattern of shared behavior[ODS02]. This is consistent with the emergent cognition paradigm discussed above, especially the co-constructed nature of the interaction, inspired by concepts of autopoiesis and structural coupling [MV80] (see Figure 2). Such mutually constructed patterns of complementary behaviour is also emphasized in Clark's notion of joint action[Cla94]. Thus, explicit meaning is not necessary for anything to be communicated in an interaction, it is simply important that the agents are mutually engaged in a sequence of actions. Meaning emerges through shared consensual experience mediated by interation. The RobotCub research programme is based on this foundational principle of interaction.

The developmental progress of imitation follows tightly that of the development of other interactive and communicative skills, such as joint attention, turn taking and language[NGPR99, Spe89, TKF99]. Imitation is one of the key stages in the development of more advanced cognitive capabilities.
If development is such an important part of cognitive systems, what is it that drives the development process? What factors motivate development? In other words, how do you exploit the phylogenetic stereotyped actions to drive the ontogenic development by which ever-richer cognitive capabilities emerge, consolidate, and, in turn, self-amplify to produce an artificial embodied agent with the understanding and communication abilities? In RobotCub, we borrow heavily from both the neurosciences and developmental psychology to guide us in identifying the necessary phyologeny, the progression of ontogenic development, the balance between the two, and the factors that drive the ontogenic development. We consider these in detail in the next section.

[4] Although communication in general, and especially language-based communication, is extremely important in the development of prospective cognition with long time horizons, such as those involved in deliberation and reasoning, to limit the extent of our research programme, we restrict ourselves to gestural communication in the RobotCub project.

Before we proceed, we must make one final comment about the role of system morphology in cognition. In both processes of co-determination and co-development, the morphology of the cognitive system not only matters and influences what developments can occur and how they occur, it is also a constitutive part of the self-organization and the structural coupling with the environment. That is, the morphology is crucial both for the systems phylogeny *and* the system ontogeny. In both instances, the morphology need not be static (and it is probably essential that it isn't static) but that it be plastic and capable of development or change. We will return to this issue of the co-development of cognition and morphology in Section 8.

Part II

The Phylogeny and Ontogeny of Natural Cognitive Systems

4 Action as the Organizing Principle in Cognitive Behaviour

Converging evidence from many different fields of research, including psychology and neuroscience, suggests that the movements of biological organisms are organized as actions and not reactions. While reactions are elicited by earlier events, actions are initiated by a motivated subject, defined by goals, and guided by prospective information.

Actions are initiated by a motivated subject. The motives may be internally produced or externally inspired but without them there will be no actions. Earlier events and stimuli in the surrounding may provide information and motives for actions, but they do not just elicit the movements like reflexes do, not even in the newborn infant. Converging evidence shows that most neonatal behaviours are prospective and flexible goal-directed actions. This is not surprising. Sophisticated pre-structuring of actions at birth is the rule rather than the exception in biological organisms.

Actions are organized by goals and not by the trajectories they form. A reach, for instance, can be executed in an infinite number of ways. It is still defined as the same action, however, if the goal remains the same. When performing movements or observing someone else performing them, subjects fixate goals and sub-goals of the movements [J*et al.*01]. However, this is only done if an action is implied: when showing the same movements without the context of an agent, subjects fixated the motion instead of the goals [FJ03]. Thus, the goal state is already represented when actions are planned [Joh00]. Evidence from neuroscience shows that the brain represents movements in terms of actions even at the level of neural processes. A specific set of neurons, 'mirror neurons', are activated when perceiving as well as when performing an action [CR04]. These neurons are specific to the goal of actions and not to the mechanics of executing them [U*et al.*01].

Actions are guided by prospective information. Adaptive behaviour has to deal with the fact that events precede the feedback signals about them. In biological systems, the delays in the control pathways may be substantial. The total delays for visuo-motor control, for instance, are at least 200-250 ms. Relying on feedback is therefore non-adaptive. The only way to overcome this problem is to anticipate what is going to happen next and use that information to control ones behaviour. Most events in the outside world do not wait for us to act. Interacting with them require us to move to specific places at specific times while being prepared to do specific things. This entails foreseeing the ongoing stream of events in the world as well as the unfolding of our own actions.

Predictive control is possible because events in the world are governed by rules and regularities. The most general ones are the laws of nature. Inertia and gravity for instance apply to all mechanical motions and determine how they will evolve. Other rules are more task specific, like those that enable us to drive a car or ride a bike. Finally, there are socially determined rules that we have agreed upon to facilitate social behaviour and to enable us to communicate and exchange information with each other. Information for predictive control of behaviour is available through both perception and cognition. Perception provides us with direct information about what is going to happen next. Our knowledge of the rules and regularities of events enable us to go beyond perception and predict what is going to happen over longer periods of time. Together the sensory based and the knowledge based modes of prospective control supplement each other in making smooth and skilful actions possible. The ultimate function of cognition is to guide actions. In adult humans, the cognitive processes involved may sometimes appear rather remotely and indirectly related to action, but it is important to point out that expressions of language are actions in their own right. In young children, the connection between action and cognition is much more direct. In the prelingual child, cognition can only be expressed through movements of the child.

Perception and action are mutually dependent. Together they form adaptive systems. No action, however prescribed, can be implemented in the absence of perception [Ber67]. Perception is needed both for planning actions and for guiding them toward their goals. However, not only does action rely on perception, it is also a necessary part of the perceptual process. For instance, active touch is required to haptically perceive the form of an object [Gib66]. The hand must move over the object and feel its form, its bumps and its indentations. The clearest example of the necessity of action for functional perception is vision itself. Our visual field consists of a very small fovea surrounded by a large peripheral visual field over which acuity rapidly deteriorates with increasing angular eccentricity. In spite of this, we have the illusion that we see equally clearly over our whole field of vision. A simple experiment shows that this is wrong. If one firmly fixates a word in a text it is hardly possible to even read the neighboring words. The illusion of an equally clear visual field is created by the fact that we move the fovea to every single detail that we want to inspect, by doing this we can inspect it with optimal resolution. The same principles hold for all modes of perceiving. Perception is always characterized by exploratory activities such as looking, listening, sniffing, tasting, and feeling [Gib66]. It is equally true that all actions also have perceptual functions. Locomotion reveals the layout of the environment, manipulation reveals object properties, and social interaction is essential for person perception. One’s movements also reveal information about the biomechanics of the body, the forces acting on it and how these change over the execution of a movement. Thus, by necessity, any action also involves perceptual actions.

In traditional terminology a distinction is made between planned movements controlled by feedforward information and unplanned movements controlled by feedback information from the movement itself. But feedback and feed-forward are deceptive concepts. Time is irreversible and what has been accomplished is only of interest for the ability to control the next part of the action. Therefore, the question is not whether a movement is controlled by feedback or feed-forward, but rather how far into the future it reaches. The development of skill is both a question of building procedures for structuring actions far ahead in time and procedures for extracting the right kind of information for the detailed monitoring of actions.

5 Prenatal Development

An organism cannot develop without some built-in ability. If all abilities are built in, however, then the organism does not develop either. There is an optimal level for how much phylogeny should provide and how much should be acquired during the life time. Most of our early abilities have some kind of built-in base. It shows up in the morphology of the body, the design of the sensory-motor system, and in the basic abilities to perceive and conceive of the world. One of the greatest challenges of development is to find out what those core abilities are and how they interact with development in building basic skills.

5.1 Morphological Pre-structuring

The most obvious way in which the child has been prepared for action is the design of its body. It is clear that hands are made for grasping and manipulating objects, feet are made for walking, and eyes are made for looking. However, there is no grand plan for evolution. It just optimizes what is at hand. Therefore the same body-part may look rather different in different species depending on its function. For instance, the limbs of horses, lions, and humans differ for obvious functional reasons. It is also true that different body parts may have evolved to serve the same function. The trunk of elephants and hands of humans are both examples of how the morphology of the body has been altered in special ways in order to facilitate object manipulation.

What is less obvious but equally true is that each of these body parts is a part of a perception-action system that also includes specially designed perceptual and neural mechanisms. The design of the body of any animal, its sensory and perceptual system, its effector system, and indeed its neural system have been tailored to each other for solving specific action problems. The changes in the morphology of the body also include adjustments of the perceptual system to improve extraction of information for controlling specific actions. For instance, the frontal positions of the eyes in primates give access to better information for controlling manual movements. It should be noted, however, that the same evolutionary change decreases the size of the visual field and decreases the ability to quickly detect predators. Precise manipulation is greatly facilitated by the evolution of detailed foveal vision, by the ability to precisely converge and accommodate the eyes on the point of interest and track objects over space, and by the evolution of direct cortico-motor-neuronal pathways that makes it possible to control individual finger movements [Kuy73].

In lower vertebrates, it often appears as if action systems have evolved independent of each other. Thus the frog seems to possess independent perceptual mechanisms for extracting spatial information needed for catching flies and for negotiating barriers [Roz76]. In higher vertebrates, movement patterns are more flexible and the perceptual skills more versatile. When a new skill evolves, the animal may re-use some of the mechanisms already evolved for other tasks instead of developing completely new ones. This leads to more general mechanisms and more generalized skills. A similar trend seems to be going on in ontogeny. The earliest appearing skills seem more task specific than those appearing later.

5.2 Pre-structuring of the Motor System

Simply providing the hardware is not sufficient for establishing a perception-action system. In addition there need to be some initial constraints on the movements produced in order to reduce the many degrees of freedom of the motor system [Ber67]. To facilitate control, the activation of muscles is therefore organized into functional synergies at the beginning of life. Synergies have both facilitating and constraining effects. For instance, the arm and finger movements of newborn infants are organized into extension and flexion synergies that make the arm and the fingers extend and flex together. These synergies simplify the control problem and enable newborn infants to direct movements of their arms in space. However, it prevents the neonate from grasping an object reached for because that would require them to flex the hand around the object while the arm is extended.

Organized movements of the human child are observable from the 9th week of gestation [dVVP82]. Within a month the foetus will begin to make organized breathing movements, open and close the mouth, yawn, suck, and swallow. They will move their arms and hands and turn the head in an organized way. There is evidence, that the fetus moves its hands and legs to touch the walls of the amniotic sack, grasp the umbilical cord, and put the thumb in the mouth. At 22 but not 18 weeks of gestation the hand movements of a foetus are planned in the sense that those directed to the eye are more smooth, decelerated, than to the ones towards the mouth [ZBD+07]. A newborn child will perform walking movements under certain conditions. This neonatal walking is organized in a similar way as in other mammals with the toes being lowered ahead of the heels [For85]. Neonatal stepping has similar frequencies for touch as when an optic flow is presented visually [BAD+]. When awake 3-day-old human infants are vertically positioned above an optic flow their stepping is related to the characteristics of the flow. It is concluded that the visual information of flow direction and velocity influences the leg movements. Other studies have shown that the stepping is influenced by the external condition in which it is performed. When infants' legs were loaded with small weights to simulate normal gains in leg fat, previously stepping infants stopped stepping [TFR84]. Conversely, when infants' legs were submerged in a tank of water to alleviate the effects of gravity, non-stepping infants stepped once again.

This innate stepping performance forms a base that many studies have ranked to have high relevance for the later walking pattern. The developmental process is intricately interwoven with the core motor abilities and already at birth infants have experience that might be crucial to their development. All these activities might be of importance for the structuring of the motor system.

5.3 Pre-structuring of the Perceptual System

Perception also requires some structuring to begin with in order to provide the necessary guidance for action. Infants must be able to perceive speech sounds in order to be ready to produce them. Research in this area has shown that speech perception actually develops ahead of speech production [Men83]. Before vision can guide looking, the visual field must be directionally structured and before it can guide object directed action, it must be able to divide up the perceptual field into object defining entities. Although little is known about when these processes of perceptual structuring start to emerge in development, some of the actions performed by newborn infants indicate that object perception is present at birth.

The early structuring of vision is accomplished prenatally and provides a beautiful example of the parsimony of the embryo-genetic process. It may serve as an example of the more general principles of neural mapping. It is a two-stage process. Both stages of mapping are necessary [vdMS88]. The first stage is primarily determined by the genotype and the second stage by the activity of the fetus. First, an abundance of axons originating at the retinal level migrate to the thalamus (the lateral geniculate nucleus) and the superior colliculus under guidance of genetically determined chemical gradients where they will form topographies crudely corresponding to the retinal topography [RH96]. The resulting projections are, however, too fuzzy for extracting specific information about the world.

At the second stage of the mapping, structured activity at the retinal level will cause connections to be modulated through competitive interactions [vdMS88]. Strong connections become strengthened [Heb49] and will successfully compete with the weaker connections for the limited synapse space available. This will transform the initial crude mapping into a detailed one. Spontaneous neural activity at the retinal level ensures that enough structured activity at the retinal is provided to map up the visual system [Sha92]. It is possible that the spontaneous activities of the foetus facilitate the mapping of the visual system. Moving the arms in front of the eyes in the womb produces moving shadows over the eyes that might assist in the mapping of the visual system. In addition, the change in the light level when the arms move in front of the eyes provides information about the contingencies between arm movements and visual input.

All sensory systems are available from birth and can be used to guide basic forms of actions. Most of them have been available in the womb and the child has had opportunities to use them. The sensory that has been least exercised is the visual system because the light that reaches the eyes is only minimally structured. At birth the visual acuity is only 3-5% of the adult one. However, this enables the child to see their hands and the gross features of another persons face.

Although perception and action are mutually dependent, there is an asymmetry between them. Perception is necessary for controlling actions and every action requires specific information for its control. Without perception there will be no action. Action is a necessary part of perceiving but only in a general sense. Specific actions are not required for producing specific percepts and action does not tell perception what to perceive. It only provides opportunities for perceiving and guides the perceptual system to where the information is.

This has clear consequences for development. The ability to extract the necessary information must be there before actions can be organized. Only then can the infant learn to control the dynamics of their motor system and gear it to the appropriate information. Take, for instance, the speech system where infants' ability to perceive the phonemic and prosodic structure of speech develops much ahead of their ability to produced those sound qualities. The infant is still able to produce sounds and show joy in doing that but the sounds have a much simpler cyclical structure than suggested by their perceptual abilities.

5.4 Forming Functional Systems

The various constraints set up by phylogeny will selectively sponsor the growth and structuring of pathways in the nervous system that are parts of functional systems which the child needs at birth [Ano64]. As a consequence of this selective, accelerated growth, neonates are prepared to sustain life in their new environment and to explore and adapt to it. Anokhin [Ano64] gives a number of examples of such accelerated growth. For instance, although the facial nerve is an isolated structure, it shows a marked disproportionate maturation of several fibers at birth. The fibers projecting to M. orbicularis oris, providing the most important movement in sucking, are already myelinated and the contacts with the muscle fibers established at a stage when no other facial muscles have such marked organization. Similar accelerated growth can be observed in the medulla oblongata. The parts related to the functional system of sucking are ready to be used, while, for instance the parts that are the source of the frontal branches of the N. Facialis, are just beginning to differentiate. The fact that the morphogenesis of the nervous system primarily follows functional rules rather than structural ones was called "the principle of systemogenesis" by Anokhin [Ano64].

6 Core Abilities

To facilitate the acquisition of particular kinds of ecologically important knowledge, basic aspects of them are prestructured in human infants. This is valid for the perception of objects and the way they move, the perception of geometric relationships and numerosities, and the understanding persons and their actions. Work with other animal species indicates that these systems have a long evolutionary history. Nevertheless, core knowledge systems are limited in a number of ways: They are domain specific (each system represents only a small subset of the things and events that infants perceive), task specific (each system functions to solve a limited set problems), and encapsulated (each system operates with a fair degree of independence from other cognitive systems) [Spe00]. Knowledge about objects, space, numbers, and people are a few of them.

6.1 Objects

A basic requirement for perceiving and interacting with the surrounding world is that it can be divided up into relatively independent units with inner unity and outer boundaries that can be handled and interacted with, *i.e.* objects.

Object perception does accord with principles governing the motions of material bodies: Infants divide perceptual arrays into units that move together, that move separately from one another, that tend to maintain their size and shape over motion, and that tend to act upon each other only on contact. To be perceived as an object, there must be well-defined and persistent outer boundaries. A heap of sand, for instance, is not perceived as an object. These findings suggest that a general representation of object unity and boundaries is interposed between representations of surfaces and representations of objects of familiar kinds [Spe90].

Perceived objects move on continuous and un-obstructed paths. When motion carries an object fully out of view, the object is expected to continue on the same path. Baillargeon and associates [BG87, AB02] habituated infants to a tall and a short rabbit moving behind a solid screen. This screen was then replaced by one with a gap in the top. The tall rabbit should have appeared in the gap but did not. Five-and-a-half-, 3.5-, and 2.5-month-old infants looked longer at the tall rabbit event suggesting that infants had detected a discrepancy between the expected and the actual motion of the rabbit in that display. When infants visually track an object that disappears temporarily behind another one during its motion they stop at the border of disappearance and shift gaze to a position at the extension of the previous trajectory just before the object reappears there [vHFS00, RvH04, KG06]. This behavior emerges around 3 months of age (RvH04) and at 4 months it is functionally mature. Then, infants will adust the latency of moving gaze to the reappearance edge to the velocity of the moving object and the width of the occluder. These behaviors are not rigid, however. If the object does not reappear at the expected location, infants quickly learn a new reappearance location [vHFS00, KG06]. Kochukhova and Gredebäck (op.cit.) found that 6-month-old infants who visually tracked a moving object that disappeared behind an occluder after having moved on a straight path began to expect the object to reappear on a path perpendicular to the original one after this had occurred on only 2 trials When the object disappears, infants do not shift gaze to the expected reappearance position right away. They rather wait until the object is about to reappear before making a saccade over the occluder [GvHB02, vHKR06]. Such behavior is seen consistently from 3-4 months of age [RvH04, vHKR06].

6.2 Numbers

Young infants have two core knowledge systems related to numbers: one that deals with small, exact numbers of objects and one that deals with approximate numerosities of sets [Spe00, FDS04]. The knowledge about exact numbers seems to have a limit of 3. Infants' discriminate 1 vs. 2 and 2 vs. 3 reliably but not any higher numbers. The exact number concept is not dependent on modality. Infants prefer to look at an array of objects that corresponds in number to a sequence of sounds. Three tones and 3 objects are perceived as equal in this respect. [SSG90]. Infants also have the ability to add these small numbers. Wynn [Wyn92] found that when one doll was hidden behind an occluder and another doll was hidden there as well, infants expected two dolls to be present when the occluder was removed. Thus, the exact core number concept seems to have a limit of 3. When 10- and 12-month-old infants were shown crackers being hidden in two different buckets, they choose the one with more crackers up to 3. With any higher numbers the choice was random [F*et al.*02]. When 14-month-old infants saw objects being hidden sequentially in a box and then were able to search for them, they retrieved all of them if the number of objects were 1, 2, or 3. However, when 4 objects were hidden, infants retrieved one of them and then stopped searching [FDS04]. The approximate number system enables infants to discriminate larger sets of entities. Xu and Spelke [XS00] found that 6-month-old infants' discriminated numerosities 8 vs. 16 using a habituation paradigm. Infants' numerical discriminations are imprecise and subject to a ratio limit: 6-month-old infants successfully discriminate 8 vs. 16 but fail with 8 vs. 12. Second, numerical discrimination increases in precision over development and adults can discriminate ratios as close as 7:8 [BKS03].

6.3 Space

Research on animals, including humans, suggest that navigation is based on representations that are dynamic rather than enduring, egocentric rather than geocentric, and and limited to a restricted subset of environmental information. Uniquely human forms of navigation build on these representations [WS02]. The evidence comes from studies of path integration, place recognition, and reorienting based on congruence finding on representations of the shape of the surface layout. Path integration has been found to be one of the primary forms of navigation in insects (see *e.g.* [MW88], birds (see e.g. [R*et al.*95], and mammals [Gal90]. Like other animals, humans can return to the origin of a path and travel to familiar locations along novel paths [L*et al.*84]. When asked to point to objects in familiar locations while moving around blindfolded, the errors made by subjects accumulate just like they do with path integration in animals. If, however, the subjects were shown just one single beam of light the errors stay small and constant. This shows that the errors were not caused by forgetting but rather by disorientation. Like animals, humans orient by recognizing places rather than by forming global representations of scenes. Gillner and Mallot [GM98] studied how people learn to navigate through a virtual neighborhood of interconnecting streets furnished with multiple landmarks. Patterns of travel provided evidence that people learn to turn in specific directions at particular places, and that their turning decisions depend on local, view-dependent representations of landmarks. That children use such a geometry-based reorientation system is also suggested by Hermer and Spelke [HS96]. They studied 1.5- to 2-year-old children who saw a toy hidden in one corner of a rectangular chamber, were then disoriented by turning, and finally released and encouraged to find the toy. In different experiments, the location of the toy was specified by the distinctive color of a single wall or by the presence of a distinctive landmark object. Like rats, children searched reliably and equally at the correct corner and at the geometrically equivalent opposite corner. Their successful use of room geometry showed that they were motivated to perform the task, remembered the object's location, and, like rats, reoriented in accordance with the shape of the surface layout but not by non-geometric landmarks.

6.4 People

An important part of core knowledge has to do with people. Infants are attracted by other people, endowed with abilities to recognize them and their expressions, communicate with them, and perceive the goal-directedness of their actions. The motions produced by a moving person are preferred over other motions in young infants. Fox and Daniel [FD82] demonstrated that 8-week-old infants preferred a point-light walker over dynamic noise or the same configuration inverted in the image plane. Intentions and emotions are displayed by elaborate and specific movements, gestures, and sounds which become important to perceive and control. Some of these abilities are already present in newborn infants and reflect their preparedness for social interaction. Neonates are very attracted by people, especially to the sounds, movements, and features of the human face [Mau85, FCSJ02]. They have a greater tendency to visually track a schematic face than one where the facial parts are scrambled inside the outer contour [JM91]. They look longer at a face that directs the eyes straight at them than at one that looks to the side [FCSJ02]. They also engage in some social interaction and turn-taking that among other things is expressed in their imitation of facial gestures [MM77]. Finally, they perceive and communicate emotions such as pain, hunger and disgust through their innate display systems [Wol87]. These innate dispositions give social interaction a flying start and open up a window for the learning of the more intricate regularities of human social behavior. Parents show a remarkable talent for responding to the infant's signals and turning them into sophisticated forms of social interaction. Rochat and Striano [SR99] suggested that this "propensity to express empathy through the echoing of affects and feelings in highly scaffolding ways is part of normal parenting and the primary source of intersubjectivity".

Spelke [Spe00, Spe03] suggests that the core knowledge systems found in infants contribute to later cognitive functioning in two ways. First, core systems continue to exist in older children and adults, giving rise to domain-specific, task-specific, and encapsulated representations like those found in infants. Second, core knowledge systems serve as building blocks for the development of new cognitive skills. When children or adults develop new abilities to use tools, to perform symbolic arithmetic calculations, to read, to navigate by maps and landmarks, or to reason about other people's mental states, they do so in large part by assembling in new ways the representations delivered by their core knowledge systems. Language presumable plays an important role in this process.

6.5 Core Motives

The development of an autonomic organism is crucially dependent on motives. They define the goals of actions and provide the energy for getting there. The two most important motives that drive actions and thus development are social and explorative. They both function from birth and provide the driving force for action throughout life.

The social motive puts the subject in a broader context of other humans that provide comfort, security, and satisfaction. From these others, the subject can learn new skills, find out new things about the world, and exchange information through communication. The social motive is so important that it has even been suggested that without it a person will stop developing altogether. The social motive is expressed from birth in the tendency to fixate social stimuli, imitate basic gestures, and engage in social interaction.

There are at least two exploratory motives. The first one has to do with finding out about the surrounding world. New and interesting objects (regularities) and events attract infants' visual attention, but after a few exposures they are not attracted any more. This fact has given rise to a much used paradigm for the investigation of infant perception, the habituation method. An object or event is presented repeatedly to subjects. When they have decreased their looking below a certain criterion, a new object or event is shown. If the infants discover the change, they will become interested in looking at the display again.

The second exploratory motive has to do with finding out about one's own action capabilities. For example, before infants master reaching, they spend hours and hours trying to get the hand to an object in spite of the fact that they will fail, at least to begin with. For the same reason, children abandon established patterns of behaviour in favour of new ones. For instance, infants stubbornly try to walk at an age when they can locomote much more efficiently by crawling. In these examples there is no external reward. It is as if the infants knew that sometime in the future they would be much better off if they could master the new activities. The direct motives are, of course, different. It seems that expanding one's action capabilities is extremely rewarding in itself. When new possibilities open up as a result of, for example, the establishment of new neuronal pathways, improved perception, or biomechanical changes, children are eager to explore them. At the same time, they are eager to explore what the objects and events in their surrounding afford in terms of their new modes of action [GP00]. The pleasure of moving makes the child less focused on what is to be achieved and more on its movement possibilities. It makes the child try many different procedures and introduces necessary variability into the learning process.

7 Development

Although all our basic behaviours are deeply rooted in phylogeny, they would be of little use if they did not develop. Core abilities are not fixed and rigid mechanisms but are there to facilitate development and the flexible adaptation to many different environments. Development is the result of a process with two foci, one in the central nervous system and one in the subject's dynamic interactions with the environment. The brain undoubtedly has its own dynamics that makes neurons proliferate, migrate and differentiate in certain ways and at certain times. However, the emerging action capabilities are also crucially shaped by the subject's interactions with the environment. Without such interaction there would be no functional brain. Perception, cognition and motivation develop at the interface between neural processes and actions. They are a function of both these things and arise from the dynamic interaction between the brain, the body and the outside world. A further important developmental factor is the biomechanics of the body: perception, cognition and motivation are all embodied and subject to biomechanical constraints. Those constraints change dramatically with age, and both affect and are affected by the developing brain and by the way actions are performed. The nervous system develops in a most dramatic way over the first few months of postnatal life. During this period, there is a massive synaptogenesis of the cerebral cortex and the cerebellum [Hut90, HD97]. Once a critical mass of connections is established, a self-organizing process begins that result in new forms of perception, action and cognition. The emergence of new forms of action always relies on multiple developments [TS03]. The onset of functional reaching depends, for instance, on differentiated control of the arm and hand, the emergence of improved postural control, precise perception of depth through binocular disparity, perception of motion, control of smooth eye tracking, the development of muscles strong enough to control reaching movements, and a motivation to reach.

7.1 Acquiring Predictive Control

If mastery of actions relies on the perception and knowledge of upcoming events, then the development of actions has to do with acquiring systems for handling such information. It has to do with anticipating both one's own posture and movements, and future events in the world. For every mode of action that develops, new prospective problems of movement construction arise and it takes time to acquire ways to solve them. The knowledge gathered through systematic exploration of a task is structured into a frame of reference for action that makes planning possible. This is the basis of skill. The importance of practice and repetition is not to stamp in patterns of movement or achieve an immutable program, but rather to encourage the functional organization of action systems [Ree96].

7.2 The Development of Perception

Two processes of perceptual development can be distinguished. The first one is a spontaneous perceptual learning process that has to do with the detection of structure in the sensory flow. As long as there is variability and change in the sensory flow, the perceptual system will spontaneously learn to detect structure and differentiate invariants in that flow that correspond to relatively stable and predictible properties of the world. The second process is one of selecting information relevant for guiding action. Infants must already have detected that structure in the sensory flow before it can be selected to guide action. It could not be the reverse. In other words, perception is not encapsulated in the actions to start with as Piaget suggested [Pia53, Pia54]. It may actually be the other way around.

7.3 Visual Development

The retina is rather immature at birth. The receptors are inefficient and only absorb a small fraction of the light that reaches the eye. Consequently, the acuity is low, only about a 40th to a 30th of the adult acuity. The discrimination of contrast is deficient to a corresponding degree. The rods and cones are evenly spread over the retina [BB88] and the cones are undeveloped. Therefore both acuity and contrast sensitivity is bad (about 2.5 - 3.5 % of the adult's acuity) and colour is poorly discriminated. These conditions change dramatically after birth. First, the cones migrate towards the fovea resulting in the massive concentration of cones in that part of the retina in adults. The rods, however, do not change position. They remain evenly distributed over the retina over development. The change in receptor distribution rules out the possibility that the infant has an innate sensitivity for certain retinal patterns or templates or that certain retinal patterns are learnt shortly after birth because the pattern of excitations will not be the same over development. As a result of the changes occurring on the retina and in the ganglions further back, the visual acuity improves dramatically during the first few months of life. At 5 months of age the acuity is adult-like.

7.3.1 Space Perception

Several of the basic cortical visual functions are not available at birth but mature during the first half year of life. Thus, colour perception is deficient at birth but functions from about 1 month of age. Certain aspects of motion perception are available at birth and are then processed subcortically. However, neonates cannot process motion direction and cannot do smooth pursuit. Both of these functions rely on functional cortical processing of motion. Von Hofsten and Rosander [vHR96, vHR97] recorded eye and head movements in unrestrained 1- to 5-month-old infants as they tracked a happy face moving sinusoidally back and forth in front of them. They found that the improvement in smooth pursuit tracking was very rapid and consistent between individual subjects. Smooth pursuit starts to improve around 6 weeks of age and attain adult levels from around 14 weeks. The ability to discriminate motion direction emerges during the same period (see [Atk00]. ERP studies show that the MT-MST area is engaged in motion processing at least from around 8 weeks of age and is fully functional from about 14-18 weeks [RGNvH06].

Visual space perception relies primarily on binocular information, motion information, and a whole set of monocular cues that induce depth in pictures. They all develop during the first year of life but at different schedules [KA98]. Let's first consider motion as information for depth. There is such information in the expansion of the retinal projection of an approaching object, the motion parallax on the retina when the subject moves, and the accretion-deletion of object structure at the edge on an occluding object when one object moves behind another. The earliest signs of sensitivity to space from motion comes from studies of looming. Reliable effects of increased blinking to approaching displays have been found in several studies with infants from less than a month on [YPL79, Nan88]. Kayed & van der Meer [KdM00] found that the youngest infants blinked when the virtual object reached a threshold visual angle, while older ones geared their blinks to the virtual object's time-to-collision. The shift did not occur until at around 6 months of age. This indicates that although young infants perceive that an object is approaching, they cannot evaluate so well when it is going to hit them.

Sensitivity to motion parallax was demonstrated in 3-month-old infants by von Hofsten, Kellman, and Putaansuu [vHKP92]. They showed infants an array of 3 vertical rods in a horizontal row, perpendicular to the line of sight. When the infant moved laterally in front of these rods, the middle one moved in phase with the infant. Afterwards they were tested with 3 stationary rods with the middle one either aligned with the other ones (as in the original display) or displaced backwards to an extent corresponding to the contingent motion. When the velocity of the contingent rod was 0.32°/s (visual angle), the infants looked significantly more at the 3 aligned stationary rods than the display where the middle rod was displaced backwards to an extent corresponding to the contingent velocity. When the contingent velocity was decreased to 0.16°/s, the looking at the test display did not show any preference. The results are consistent with the idea that young infants utilize small contingent optical changes as information about depth. The results do not uniquely imply this interpretation, however. It might simply be that infants are very sensitive to optical changes contingent on their own motion. These optical changes do appear special in that infants' sensitivity to them exceeded what has been found in other studies of motion sensitivity by almost an order of magnitude (see e.g. [AS90, DF89].

Binocular depth perception relies on two mechanisms – sensitivity to the convergence of the eyes and sensitivity to binocular disparity. Convergence gives absolute distance and disparity relative depth to objects in the surrounding. By 1 month of age, convergence operates accurately for distances beyond 20 cm ([HRGfA92]. Von Hofsten [vH82a] showed that by 5 months of age, infants use convergence information when programming reaching movements but convergence may be used much earlier in life, maybe even at birth. Kellman, von Hofsten, van der Walle & Condry [KvHvdWC90] showed displays to young infants that contained several stationary objects and one that moved contingent on the movement of the infant. In order to perceive which object was moving, the infant had to correctly perceive the distance to it. 8-week-old infants consistently disrciminated displays containing a moving object from those with only stationary ones. As 8-week-olds have been found to process binocular disparity information [F*etal*0], this is most probably responsible for the effect. Other signs of binocular depth perception have been observed in 8-week-old infants, but many infants first showing sensitivity a month or so later. Birch, Gwiazda, & Held [BGH82] found reliable preferences at 12 weeks of age for crossed disparities and at 17 weeks of age for uncrossed. Improvement in stereoscopic acuity once it appears is quite rapid. Sensitivities improved from 60' visual angle to less than 1' in just a few weeks [HBG80]. In this respect, the ability to process stereoscopic depth show a parallel development relative to the ability to process visual motion. Indicators of perceived motion show that this ability also emerges within just a few weeks [vHR97, Atk00].

The development of sensitivity to pictorial depth information comes primarily from studies by Yonas and colleagues (see [YAG87]. Many of these studies used reaching as dependent measure. They systematically examine the different depth cues, including Linear perspective, familiar size, interposition, and shading. The results are quite consistent and suggest that infants do not utilize pictorial depth cues to guide reaching until they are 6-7 months old. It is possible that several of the pictorial depth cues originate from dynamic situations. For instance interposition refers to the cue that an object that is partly hidden by another is perceived to continue behind it. Granrud & Yonas [GY84] found that 7-month-old but not 5-month-old infants utilize this cue when reaching for objects. The dynamic version of this cue is the gradual accretion and deletion of object texture as one object goes behind another. 5-month-old infants reliably use this information in predicting when and where an object that disappears behind another will reappear on the other side [BG06].

In summary, young infants primarily define objects by binocular information and relative motion. Only a few months later do infants become able to use cues like surface structure, shading, familiar size, linear perspective and interposition.

7.3.2 Object Perception

The rules by which infants perceive objects as separate entities are similar to the ones used by adults. Objects are defined by outer boundaries and inner unity that are preserved over time. To be perceived as an object, there must be well-defined and persistent outer boundaries. A heap of sand, for instance, is not perceived as an object. This suggests that a general representation of object unity and boundaries is interposed between representations of surfaces and representations of objects of familiar kinds [Spe90].

To define the outer boundaries and the inner unity, motion information is relatively more important than static information early in life. Infants divide perceptual arrays into units that move together, that move separately from one another, that tend to maintain their size and shape over motion, and that tend to act upon each other only on contact. Two units that move relative to each other are perceived as separate objects and two units that move together are perceived as a single object [vHS85], SvHK89]. Units that are separated in depth, thus creating relative retinal motion as the subject move, are also perceived as separate objects. If only parts of an object are visible and the space between them is occluded by a nearer object, the parts are still perceived as belonging to one object if the occluded object moves or the subject moves. Kellman and Spelke [KS83] found that object pieces protruding on each side of an occluder were not perceived as belonging to the same object if they were stationary. However, if the pieces moved with a common motion along the occluder , 3-month-old infants perceived them to belong together and to be connected behind the occluder. This was the case both when the pieces showed good continuation behind the occluder, such as being parts of a single rod, and when they were totally dissimilar. Smith et al. [SJS02] found that when the pieces protruding from behind the occluder were misaligned relative to each other, common motion had a somewhat weaker binding effect. They concluded that alignment information could enhance perception of object unity either by serving directly as information for unity or by optimizing the detectability of motion-carried information for unity. Van de Walle & Spelke showed 5-month-old infants objects whose center was fully occluded and whose ends were visible only in succession. Infants perceived this object as one connected whole when the ends of the object underwent a common motion but not when the ends were stationary.

Static object information such as good form, surface colour and texture similarity are much less import as determinants of object unity and boundaries in young infants. Spelke et al. [SvdW93] presented adults and infants with simple but unfamiliar displays in which texture similarity, good form, and good continuation either specified one object or two objects. Object perception was assessed by a verbal rating method in the adults and by a preferential looking method in the infants. The Gestalt relations appeared to influence the adults' perceptions strongly. However, the relations appeared to have no effect on the perceptions of 3-month-old infants and weak effects on the perceptions of 5-month-old and 9-month-old infants. This suggests that motion information dominates infants' perception of objects. Three-month-old infants group surfaces in accord with the cohesion principle [SvdW93]. Presented with an array of adjacent surfaces, they perceive a connected body that maintains its connectedness as it moves. These principles apply equally to familiar and unfamiliar forms. Developmental changes in object perception occur only slowly towards a more mature mode where the gestalt principles of good form, surface colour and texture similarity play a more important role.

Colour contributes to the identification of object at the end of the first year of life Wilcox et al., [WWC+07] found that multi-modal exploration of objects (visual and tactile), but not unimodal (visual only), exploration of objects prior to an individuation task increased 11-month-old infants sensitivity to colour differences.

7.4 The Development of Basic Modes of Action

The principles outlined above, will be exemplified with four different modes of action: posture and locomotion, looking, reaching and manipulation, and social skills.

7.4.1 Development of Posture and Locomotion

Basic orientation is a prerequisite for any other functional activity [Gib66, Ree96] and purposeful movements are not possible without it. This includes balancing the body relative to gravity and maintaining a stable orientation relative to the environment. As Reed [Ree96] states, “maintenance of posture in the real world involves much more than simply holding part of the body steady; it is maintaining a set of invariant activities while allowing other activities to vary” (p. 88). Gravity gives a basic frame of reference for such orientational stability and almost all animals have a specialized mechanism for sensing gravity (in humans it is the otoliths). In addition, vision provides excellent orientational information as does proprioception. The contribution of vision is crucial for supporting balance prospectively.

Gravity is also a potent force and when body equilibrium is disturbed, posture becomes quickly uncontrollable. Therefore, any reaction to a balance threat has to be very fast and automatic. Several reflexes have been identified that serve that purpose. For instance, when one slips, a series of fast automatic responses are elicited that serve the purpose of regaining balance. Postural reflexes, however, are insufficient to maintain continuous control of balance during action. They typically interrupt action. Disturbances to balance are better handled in a prospective way, because if the disturbance can be foreseen there is no need for an emergency reaction and ongoing actions can continue. Another threat to balance are one’s own movements. When a body part is moved, the inertia created by the movement will push the body out of equilibrium if nothing is done about it. The movement will also shift the point of equilibrium and that will also disturb balance. Therefore, the effects of one’s own movements must be foreseen and prepared for in order to maintain ongoing activity.

At around 3 months, infants show the first signs of being able to actively control gravity. When in a prone position they will lift their head and look around. To hold the head steadily, its sway must be correctly perceived and used to control head posture. Such control seems to be attained over the first few weeks of head lifting. The next step in mastering postural control is controlling the sitting posture. This is normally accomplished around age 6-7 months and requires the child to control the sway of both head and trunk in relation to each other. This could be accomplished in a large number of ways because many different muscle groups affect the sitting posture. Woollacott, Debu, and Mowatt [WDMM87] found that infants did not show a consistent postural response synergy while sitting until around 8 months of age. Hadders-Algra, Brogren, and Forssberg, [HABF96] tested 5- to 10-monthold infants' postural adjustments when sitting on a platform and subjected to slow and fast forward and backward displacements. They found that from the youngest age onwards rather variable, but direction specific muscle activation patterns were present. With increasing age the variation in muscle activation pattern decreased resulting in a selection of the most competent patterns. Barela et al. [B*etal*00] examined whether there is any developmental change in the coupling between visual information and trunk sway in infants as they acquire the sitting position. Six-, 7-, 8-, and 9-month-olds sat inside a moving room that oscillated back and forward at frequencies of 0.2 and 0.5 Hz. Relative phase showed that at 0.2 Hz, infants were swaying with no lag but at 0.5 Hz they were lagging the room. The results showed that the coupling between visual information and trunk sway in infants varies with the visual stimulus but does not change as infants acquire the sitting position.

In upright stance, the body acts as a standing pendulum. The natural sway frequency of a pendulum is inversely proportional to the square root of its length. This means that the balancing task is much more difficult for a child than for an adult. For instance, a child who is only half the size of an adult will sway with a frequency which is 40 percent higher than that of the adult and will consequently have 40 percent less time in which to react to balance disturbances. In other words, when, by the end of the first year, infants start to be able to stand independently they have mastered a balance problem more difficult than at any time later in life. Barela, Jeka & Clark [BJC99] made a very nice demonstration of the development of predictive control of standing posture. The infants simply stood and held a handrail. The forces that the subject applied to the rail and his/her sway were simultaneously recorded. Four groups of infants were studied according to their postural maturity; prestanding, standing alone, walk onset (>3 steps), and post walking (walking for more than 1.5 months). The results show that for the first 3 groups the forces applied to the rail lagged the sway but for the post walking infants, the forces preceded the sway.

Vision is quite superior in detecting small body displacements, and with it, the subject can be more efficient in using prospective control for controlling body sway. Lee and Aronsson (Lee and Aronsson, 1974) showed that infants who have just attained upright stance are quite sensitive to peripheral visual information for body displacement. They positioned standing infants in a room with movable walls and ceiling (the moving room), and when they moved these surrounding structures, the infants lost their balance in the direction predicted by the visual flow. With more experience of standing, children were not as easily overthrown by the visual flow alone. Bertenthal, Rose, and Bai [BRB97] showed that the sensitivity to visual flow improves over the months after upright stance has been achieved. Visual information is especially important for dynamic postural control, that is, when maintaining balance while moving around. Fraiberg [Fra77] found that in a sample of blind children, 90 percent were delayed past the upper limits of sighted children as given by Bayley [Bay69] when walking independently across a room.

Special demands are associated with balance control during bodily activities. In order to maintain balance during limb movements, the subject must know about the contingencies between the limb movements, the reactive forces that arise during movement, and the displacement of the point of gravity. Adults seem to counteract disturbances to the postural system in a precise way ahead of time.
Von Hofsten and Woollacott [vHW90] found such anticipatory adjustments of the trunk in 9-monthold infants reaching for an object in front of them while balancing the trunk. Witherington et al. [W*etal*02] examined the timing of activation of the gastrocnemius muscles when standing infants pulled a drawer that resisted pulling by a weight attached to it. Activation of this muscle counteracts the tendency to fall forward during pulling but only if is activated slightly ahead of the pull. Adults activate the

gastrocnemious muscles 50 ms before the arm starts pulling. Witherington et al. [W*etal*02] found that infants activated this muscle ahead of pulling to an increasing extent between 10 and 17 months of age. The emergence of independent walking coincided with marked increases in anticipatory postural adjustments of the gastrocnemius muscle relative to pull onset.

Because of its central role in movement production, postural control becomes a limiting factor in motor development. If the infant is given active postural support, goal directed reaching can be observed at an earlier age than is otherwise possible. For instance, the neonatal reaching observed by von Hofsten [vH82b] was performed by properly supported infants. For these reasons, development of reaching and other motor skills should be studied in the context of posture. However, there are only few studies that have seriously considered the influence of such contextual factors. Rochat and associates [Roc92, RG95] showed that the onset of self-sitting made infants transfer from two-handed to one-handed reaching. They suggested that this was because the newly attained posture could be easily disturbed and that two-handed reaching was more threatening to balance than one-handed. Rochat also observed that when infants who were sitting independently reached forward with one hand, the other one often moved backwards to preserve the point of equilibrium.

Before the onset of bipedal walking two types of locomotion are observed in infants: crawling and cruising. The standard crawling with knees and hands is the most common type. However, the alternatives (locomotion with slithering on the belly or sitting) are practiced more in homes that have polished floors than in homes with rugged carpets. For the later alternative the classic or standard crawling is the most common type of locomotion. Recently, it has been shown [RNR+08] that standard crawling shares most of the basic principles of other vertebrate quadruped gaits. In a study by Haehl et al. [HVU00], it is suggested that cruising represents an important transition from quadruped to bipedal locomotion. Using support the infant learns to control the trunk and consequently improving the postural control.

During the first year of independent walking, toddlers improve their gait kinematics, master the postural instability, and the pendulum mechanism of walking [IDC+05, MMC+05]. One important parameter is head control. Ledebt and Wiener-Vacher [LW96] concludes that head stabilization in space is achieved during the first weeks of independent walking, (IW). During the first year of IW, the degree of synchronization between head rotations in the pitch plane and vertical translations increases. Another parameter, reflecting balance control, is step length. New walkers have very short lengths (≈ 12 cm), and with experience these increase ahead of step velocity (25 cm, and 25 to 80 cm/s respectively) [BA08]. Mastering the ability of bipedal walking is evidently a process of both learning and development. A key question is how a changing environment as well as bodily changes will challenge the infant's control of locomotion. Berger and Adolph [BA06] writes "the ability to detect affordances lies at the heart of adaptive locomotion". They found, for example, that after 10 weeks of experience, infants geared their locomotor decisions to the possibilities for action. Ivanenko et al [IDL07] summarized different theories for neural control of motion: dynamic systems theory, neuronal group selection, growth and environment.

Recently, two studies have focused on neurophysiological and behaviour evidence for how learning takes place. Sanefuji et al [SOH08] presented crawlers or walkers with point-light displays of similar actions. They found that crawlers preferred to look at crawling infants, and walkers at walking infants. It was concluded that transformations in the sensory-motor domain may be represented similar to those in the physical-visual one, thus supporting a mirror neuron function. This is further demonstrated in a study of van Elk et al (vE+08). They measured mu-suppression in EEG for crawlers and walkers when they observed similar videos of crawlers and walkers. The result was that the observation of crawling gave more mu suppression in crawlers, and the observation of walkers induced more mu rhythm suppression in walkers.

7.4.2 Development of Looking

Although each perceptual system has its own privileged procedures for exploration, the visual system has the most specialized one. The whole purpose of movable eyes is to enable the visual system to explore the world and to stabilize gaze on objects of interest. Vision is able to maintain contact over distance. It therefore becomes extremely important in establishing and maintaining social interaction and in learning by observation (for instance, imitation). The development of oculomotor control is one the earliest appearing skills and marks a profound improvement in the competence of the young infant. It is of crucial importance for the extraction of visual information about the world, for directing attention, and for the establishment of social communication. Controlling gaze may involve both head and eye movements and is guided by at least three types of information: visual, vestibular, and proprioceptive. How do young infants gain access to these different kinds of information, how do they come to use them prospectively to control gaze, and how do they come to coordinate head and eyes to accomplish gaze control? Two kinds of task need to be mastered, moving the eyes to significant visual targets and stabilizing gaze on these targets. Each of these tasks is associated with a specific kind of eye movement. Moving the eyes to a new target is done with high speed saccadic eye movements and stabilizing them on the target is done with smooth pursuit eye movements. The second task is, in fact, the more complicated one. In order to avoid slipping away from the target it requires the system to anticipate forthcoming events. When the subject is moving relative to the target, which is almost always the case, the smooth eye movements need to anticipate those body movements in order to compensate for them correctly. When the fixated target moves, the eyes must anticipate its forthcoming motion.

Shifting Gaze

The ability to shift gaze is of crucial importance for the development of visual perception, because it turns the visual sense into an efficient instrument for exploring the world. The saccadic system for shifting gaze develops ahead of the system for smooth tracking. It is functional at birth and newborn infants are fairly skilled at moving gaze to significant events in the visual field. The development of looking requires ability to shift and maintain attention on specific objects and events. The ability to control these actions is a basic aspect of cognitive development. What infants look at reflect their cognitive development and their interests in what is happening around them. Shifting gaze is preceded by an attentional shift which involves a process of disengaging attention to the current fixated target and moving the eyes to a new target. The ability to engage and disengage attention on targets is present at birth and develops rapidly over the first half year of life. Visual attention in infants is primarily guided by the attractiveness of objects and the predictability of events. Only at preschool age do children begin to scan the surrounding in systematic ways. Then they become able to solve problems like finding the differences between two pictures.

Tracking Eye Movements

Several studies on eye movements indicate that newborn infants have only limited ability to track a moving target smoothly. Dayton and Jones [DJ64] found that neonates pursued a wide angle visual display with smooth eye movements but the eye movements became rather jerky for a ”small” target. These results were supported by several other studies [BC92, KVKD79, Asl81]. Rosander and von Hofsten [RvH00] also found that 1-month-old infants and younger tracked a large moving vertical grating in a smoother way than a small moving target. However, when the saccades were eliminated from the records the residual smooth tracking did not differ for the two targets. In other words, the reason why the tracking of a small target looks jerky is because infants make frequent catch-up saccades in an effort to be on target which they do not need with a large target. The reason is simple.
With a wide-field pattern of vertical stripes, the eyes are always on the target, however they move.

From about 6 weeks of age, the smooth part of the tracking improves rapidly. This was first observed both by Dayton and Jones [DJ64] and by Aslin [Asl81]. Von Hofsten and Rosander [vHR96, vHR97] recorded eye and head movements in unrestrained 1- to 5-month-old infants as they tracked a happy face moving sinusoidally back and forth in front of them. They found that the improvement in smooth pursuit tracking was very rapid and consistent between individual subjects. Smooth pursuit starts to improve around 6 weeks of age and attain adult levels from around 14 weeks. The effect of target velocity depended on age. At 2 months of age the proportion of smooth pursuit in the slowest condition (0.2 Hz and 10 deg. amplitude) was almost twice as high as it was in the fastest condition (0.4 Hz and 20 deg. amplitude). At 4 months of age, the proportion of smooth pursuit was high in all conditions and approached adult values.

In order to stabilize gaze on a moving object during tracking, the smooth pursuit must anticipate its motion. Two such predictive processes have been observed in adult visual tracking [Pav90]. One uses the just seen motion to predict what will happen next through a process of extrapolation. Such predictions are in accordance with inertia which presumes that a motion with a certain speed and direction will continue with the same speed and in the same direction unless it is affected by a force in which case the motion will change gradually. The extrapolation process is important in predicting object motion over small time windows but it cannot handle prediction over larger time frames. With increasing time there are growing possibilities that intervening events will alter the motion. Neither can it handle abrupt motion changes because such changes do not reveal themselves in the just seen motion. In order to investigate the development of these predictive processes, von Hofsten and Rosander [vHR97] studied visual tracking of two motion functions, one sinusoidal and one triangular. The sinusoidal motion can be predicted by extrapolation but not the triangular one. A triangular motion is characterized by constant velocity between the end points where the motion abruptly reverses.

Von Hofsten and Rosander [vHR97] found that the amplitude of head tracking increased very much between 3 and 5 months of age. At 5 months the amplitude of the head tracking was sometimes as large as the amplitude of the object motion. The problem was that the head still lagged the target at that age (1/3 sec or more). In order to stabilize gaze on the target, the eyes must then lead. This creates a phase differences between the eye and head tracking that may be so large that the eye tracking and the head tracking counteract each other. Instead of contributing to stabilizing gaze on the fixated moving object, head tracking may then deteriorate gaze stabilization. In fact, the task would be much simpler if the head had not moved at all. The reason why infants persisted in engaging the head can only be because they are internally motivated to do so. Just as in the early development of reaching this is an expression of important developmental foresight because eventually, the ability to engage the head will result in much more flexible tracking skills.

Compensatory Gaze Adjustments

Both visual and vestibular mechanisms operate to compensate for head movements unrelated to fixation. The visual one aims at stabilizing gaze on the optic array by minimizing retinal slip while the vestibular one aims at stabilizing gaze in space. The visual mechanism is designed to work at slow optical changes and its performance begins to deteriorate at frequencies above 0.6 Hz [BB78, Hy´e83]. The vestibular mechanism functions most optimally above 1 Hz where the gain approaches unity and the phase lag approaches zero [Bar93]. Head movements unrelated to visual tracking are generally faster and more dynamic than the tracking itself and the eye movements that compensate for those head movements are predominantly guided by vestibular information. This mode of control functions at birth.

7.4.3 Development of Reaching and Manipulation

Reaching

Visual control of the arm is present at birth [vH82b, vdMvdWL95, vdM*et al.*96]. Infants can also move the fingers in a differentiated way, but they cannot control them in grasping or manipulating objects. Both arm movements and finger movements are governed by global extension and flexion synergies [vH84]. When the arm extends the fingers extend too and when the fingers flex the arm also flexes. Von Hofsten [vH84] found that the hand was either open or opened during the extension of the arm in about 70 percent of the extended arm movements. The opening of the hand did not seem to be a function of the act of reaching towards the object because the same thing happened when the child extended the arm without looking at the object. This pattern was also observed in young rhesus monkeys by Lawrence and Hopkins [LH76]. They found that newborn monkeys had difficulties in grasping an object they had reached for and if they had finally closed the hand around it, they had difficulties in releasing it after they had pulled it towards them.

Von Hofsten [vH84] found that the synergistic arm-hand pattern changed dramatically at 2 months of age. The coupling was then broken, and instead of opening the hand, the child had a strong tendency to fist it during the extension of the arm. At the same time the movements became more vigorous and appeared more voluntary, as if the child really tried to attain the object [vH86]. A few weeks later, the subjects were again observed opening the hand during the extension of the arm but then only when the arm movement was visually directed toward the object. The infants then also started closing the hand when it was close to the object, suggesting that the global extension-flexion pattern had developed into a differentiated pattern where arm and hand were more independently controlled.

Reaching for stationary objects appears between 12 and 18 weeks [CMA+93] and catching moving objects appears at approximately 18 weeks (vH79, vH80). Just as infants' first eye movements are saccadic and lagging rather than smooth and on-target, their first goal-directed reaches and catches are typically jerky and crooked. The transition from pre-reaching to reaching was studied by Thelen et al. [TZ93]. They found that each infant had its own individual way of moving its arms; some moved them more slowly with rather damped movements and some more vigorously. Overall, the early reaching attempts were characterized by much variability which casts doubt on the notion that early movements are stereotyped. During the transition from prereaching to successful reaching and grasping the movements became less variable as the infants came to control the intrinsic dynamics of their arms.

Studies of reaching kinematics [vH79, vH91, Ber96] show that early reaches are rather segmented in contrast to adult reaches which consist of a single bell-shaped velocity curve. Von Hofsten [vH79] defined movement units as segments of the reach, each consisting of an acceleration and a deceleration phase. Corrections are more pronounced during faster reaches [TCS96]. Movement units and direction changes decrease after a few months until infants' reaches and catches are made up of only two movement units, the first to bring the hand near the target and the second to grasp it. With age, prospective extrapolations of target motion become less dependent on continuous visual information. By 9 months, infants reach for moving objects on an unobstructed path but inhibit reaching when a barrier blocks the path [KCS+03]. Six-month-old infants do not plan reaches for moving objects that are temporarily occluded but wait until the object has reappeared [vHFS00]. By 11 months, however, infants catch moving objects as they appear from behind an occluder (vdM+95].

Early reaching consisted of several such segments but the number decreased rapidly with increasing age. Already after a few months of experience with reaching, the number of segments approached two units, one associated with the approach and one with the grasping act. Von Hofsten [vH93] interpreted this development as reflecting increased prospectivity of the reaching action. What makes infants' initial reaches so jerky and crooked? One possibility is that movement units reflect visual corrections for a misaligned arm path. However, infants successfully reach for objects in the dark within a week or two of reaching in the light [CMA+93], suggesting that they can use proprioceptive information to guide the reach. Indeed, by 5 to 7 months, infants can catch moving objects without sight of their hand by gauging the speed of the glowing object in the dark [RBC96] and by 9 months, they preorient their hands to grasp objects in the dark [MCA+01]. Possibly, younger infants have less ability to anticipate the reactive forces that result from the movement itself [BCG+96, TCS96, vH97]. Or, infants may have little motivation for efficient reaching [Wit08]—the functional penalty for extra movement units is low—and might even use variable arm paths to explore the capabilities of their new action system (Ber96, *BCG*+96].

In the act of reaching for an object there are several problems that need to be dealt with in advance, if the encounter with the object is going to be smooth and efficient. The reaching hand needs to adjust to the orientation, form, and size of the object. The securing of the target must be timed in such a way that the hand starts to close around the target in anticipation of and not as a reaction to encountering the object. Such timing has to be planned and can only occur under visual control. Infants do this from the age they begin to successfully reach for objects around 4-5 months of age [vHR88].

From the age when infants start to reach for objects they have been found to adjust the orientation of the hand to the orientation of an elongated object reached for [LAB84, vHFZ84, vHJ05]. Von Hofsten and Johansson [vHJ05] found that, when reaching for a rotating rod, infants prepare the grasping of the object by aligning the hand to a future orientation of the rod. Adjusting the hand to the size of a target is less crucial. Instead of doing that, it would also be possible to open the hand fully during the approach that would lessen the spatial end point accuracy needed to grasp the object. Adults use this strategy when reaching for an object under time stress [WTF86]. The disadvantage is the additional time it takes to close a fully opened hand relative to a semi-opened hand. Von Hofsten and Rönnqvist [vHR88] found that 9 and 13 month-old infants, but not 5-month-olds, adjusted the opening of the hand to the size of the object reached for. They also monitored the timing of the grasps. For each reach it was determined when the distance between thumb and index finger started to diminish and when the object was encountered. It was found that all the infants studied including those that just recently had started to reach for objects successfully began to close the hand before the object was encountered. For infants of 9 months and younger the hand first moved to the vicinity of the target and then started to close around it. For the 13 month-olds, however, the grasping action typically started during the approach, well before touch. In other words, at this age grasping started to become integrated with the reach to become one continuous reach-and-grasp act.

A remarkable ability of infants to time their manual actions relative to an external event is demonstrated in early catching behavior [vH80, vH83, vHL79]. Von Hofsten and Lindhagen [vHL79] found that infants reached successfully for moving objects at the very age they began mastering reaching for stationary ones. Eighteen-week-old infants were found to catch an object moving at 30 cm/sec. Von Hofsten [vH80] found that the reaches were aimed towards the meeting point with the object and not towards the position were the object was seen at the beginning of the reach. Von Hofsten [vH83] also found that 8-month-old infants successfully caught an object moving at 120 cm/sec. The initial aiming of these reaches was within a few degrees of the meeting point with the target, and the variable timing error was only around 50 msec. The studies show that infants predict the future position of a moving object, but they tell us little about the nature or limits of these predictions. Systematic study of the principles guiding predictive reaching requires manipulation of the spatial as well as the temporal properties of object motion. This was done by von Hofsten et al. (vH*etal*98). Infants were presented with an object that moved into reaching space on four trajectories: two linear trajectories that intersected at the center of a display and two trajectories containing a sudden turn at the point of intersection. Infants' tracking and reaching provided evidence for an extrapolation of the object motion on linear paths, in accord with the principle of inertia. This tendency was remarkably resistant to counterevidence, for it was observed even after repeated presentations of an object that violated the principle of inertia by spontaneously stopping and moving on a nonlinear path.

Grasping

When grasping first emerges, infants may use one as well as both hands. The first grasps are power grasps and engage the whole hand. Soon thereafter, however, the radial part of the hand becomes increasingly important for grasping. Although grasping then still involves the whole hand, it tends to be focused on the two most radial fingers and the thumb. Newell et al. [NSM+89] studied 4-8-month-old infants as they grasped objects that varied in size and shape. The findings revealed that infants as young as 4 months systematically differentiate grip configurations as a function of the object properties in essentially the same way that 8-month-old-infants do. The difference was that younger 4-month-old infants used the haptic system in additional to the visual system for information pick-up regarding object properties, whereas 8-month-old infants predominantly used information from the visual system alone to differentiate grip configurations according to the object properties. Siddiqui [Sid95] presented 5, 7, and 9 months old infants with objects varying from 0.5 to 14.0 cm in diameter. The findings were similar to Newell et al. (op.cit.) in the sense that 5 months differentiated grip configurations as a function of object size. The number of grasps involving the two or three most radial digits (thumb, index finger, and long finger) increased greatly over this age span. At 9 months of age these kinds of grasps were 10 times more frequent than at 5 months of age. However, at each age level, when only the two or three most radial digits were used, the reaches were typically directed at the two smallest objects. From around 9-10 months of age, infants begin to grasp objects with finger movements that are relatively independent. The independent control of the fingers is made possible by the maturation of the direct cortico-moto-neuronal pathways [Kuy73]. When infants develop such finger control, they are able to grasp very small objects with just the index finger and the thumb in precision grasping.

From the age when infants start to reach for objects they have been found to adjust the orientation of the hand to the orientation of an elongated object reached for [LAB84; vHF84, vHJ06]. The adjustments are crude to begin with but become more precise with age. However, they are never complete. Around 10-15 deg. are always left to be adjusted after contact. When attempting to catch a rotating rod, [vHJ06] found that infants prepare the grasping of the object by adjusting the hand to a future orientation of the rod. As they approached the rotating rod from any starting position, they rotated the hand with the rod. Adjusting the hand to the size of a target is less crucial. Instead of doing that, it would also be possible to open the hand fully during the approach which would lessen the spatial end point accuracy needed to grasp the object. Adults use this strategy when reaching for an object under time stress [WTF86]. The disadvantage is the additional time it takes to close a fully opened hand relative to a semi-opened hand. Von Hofsten and Rönnqvist [vHR88] found that 9 and 13 month-old infants, but not 5-month-olds, adjusted the opening of the hand to the size of the object reached for. They also monitored the timing of the grasps. For each reach it was determined when the distance between thumb and index finger started to diminish and when the object was encountered. It was found that all the infants including those that just recently had started to reach for objects successfully started to close the hand before the object was encountered. For infants of 9 months and younger the hand started to close rather late during the approach but well before touching the object. For the 13 month-olds, however, the closing of the hand typically started in the middle of the approach. In other words, the hand opened up during the first half of the approach and closed during the second half. Thus, at this age grasping started to become integrated with the reach into one continuous reach-and-grasp act.

An object is optimally grasped over an opposition space that goes through the centre of mass of the object. To investigate infants' tendency to grasp objects in this way, Barrett & Needham (BN08) presented relatively large symmetrical and asymmetrical objects to 11- and 13- month-old infants. To be able to grasp these objects, infants had to use both hands. The point of contact of each hand was measured and how far the two hands were from the centre of mass of the object. It was found that at first contact, all infants grasped the asymmetrical object further from its centre of mass than the symmetrical object. In addition, results showed that the older infants were better able to correct for less stable hand placements (that is closer to the centre of the object than the centre of mass), to maintain control of the objects without dropping them.

Bimanual Coordination

There is no consensus for the definition of what constitutes a bimanual reach. According to Corbetta and Thelen [CT96], it is enough that both hands move in the approximate direction of the object to constitute a bimanual reach. Other studies require that both hands end up at the object [BN08]. Another problem has to do with how close in time the two limbs approach the object. If one hand approaches the object a second or more after the first one, it is generally agreed that the reaches should be counted as two separate ones. If the time difference is less, however, the question arises when the reaches with the two hands merged into one bimanual reach. The question is also whether the two hands have to do the same thing or can do complimentary things. An action approach defines a bimanual action as one where both hands serve the same goal.

Fagard et al [FL05] found that when grasping and manipulating objects, it seems that there is not a certain limit of object size when both hands are needed - these actions are more task dependent and can be related to object shape and orientation or to the intended action. (for instance, banging with one hand and lifting with two). Both hands are more often engaged when the child is reaching for large objects, slippery objects, and moving objects. In some tasks, the object that is being grasped, is moved between the hands, in others, one hand assumes support while the other manipulates the object.

While much effort have been devoted to how infants approach and grasp objects, very few studies have focused on the manipulation of objects after they have been grasped. Even when only one hand is used for grasping objects, two hands are most often used for manipulating them. The two hands are also engaged when the subject performs complementary actions like squeezing, tearing, and pulling. Finally, the two hands are engaged when the child performs an action involving two objects like banging one object against another. The object may then be transported from one hand to the other and back again several times while it is being rotated. It is obvious that the function of such manipulations is to inspect the object from many different angles. Other manipulations that involve both hands are stretching, tearing and wrinkling papers, crumbling bread, and bending and squeezing elastic objects. Infants are engaged in such actions from the time they master reaching for and grasping object at around 4 months of age. One hand is most often used when the object, for instance, is banged against a surface. For infants aged between 6 and 36 months Fagard and Lockman [FL05] studied the use of one or both hands in different conditions: simple grasping, precision grasping, grasping with bimanual manipulation and object exploration. They found that there was a strong decrease in bimanual grasping between 30-36 and 48 months. Increasing the precision required for grasping decreased the variability of the grasping patterns and increased the frequency of right-handed strategies. In contrast, grasping of objects affording various explorations and subsequent exploratory behaviors were even less clearly lateralized than simple grasping. In an object-exploratory task, bimanual use dominates. Exploration was mostly visual-manual exploration in all ages. For banging one hand was used, mouthing both although these behaviours were considerably less frequent.

Recent studies indicate that laterality doesn't just mature. It is not very stable during the first year of life but rather dependent on the task performed by the infant [Ram85]. If children grasp objects far to the left, the left hand is predominantly used and when they grasp objects far to the right, the right hand is predominantly used. Laterality also has to do with the roles assumed by the two hands. For instance, when opening a jar, the left hand may hold the jar while the right hand unscrews the lid.

Fagard, Spelke, & von Hofsten [FSvH08] investigated hand preference, midline crossing, hand cooperation, and visual-field asymmetry in 6-, 8-, and 10-month-old infants who reached for and grasped a moving object by comparing how performance depended on the object's direction of motion (from right to left versus left to right). The object moved on a large circular trajectory in the horizontal plane. The results show that 6-month-olds reached for the object with the ipsilateral hand (from where the object arrived) and grasped it with the contralateral hand. The grasping, but not the reaching, showed a right-hand bias. In the 8-month-olds, the ipsilateral reaching and contralateral grasping was overshadowed by a strong right-hand bias. Finally, the 10-month-olds both reached and grasped preferentially with their ipsilateral hand or with both hands, especially when the object arrived from the left. These age-related changes in reaching strategies seem to be associated with an increase with hand preference coupled with improved manual skills. They support the hypothesis that laterality is more pronounced in a demanding task. The task is difficult for the 6-month-olds and they have not developed very strong hand preference. It is also difficult for the 8-month-olds, who master the task, and with strong expression of laterality. The mastery of the 10-month-olds is more relaxed with weaker laterality.

The results do not support the hypothesis that maturation of manual skills is associated with stronger tendency to cross the midline. On the contrary, Fagard et al. [FSvH08] found that midline crossing was most common in the youngest infants and least common in the oldest ones. The results indicate that, in addition to the need for predicting the path of a moving object, motor constraints due to spatial compatibility, hand preference and bimanual coordination must be taken into account to understand age differences in grasping a moving object.

Manipulation

The close connection between vision and manipulation makes it also possible to learn about object affordances by viewing events that engage them and other people manipulating them. This is especially relevant when learning about the functions of tools. Lockman [Loc00] suggested that tool used may be a more continuous development than previously believed and that it is rooted in in the perception-action routines that infants employ to gain knowledge about their environments. He suggested that in order to learn more about tool use development, research should focus on the processes by which children detect and relate affordances between objects, coordinate spatial frames of reference, and incorporate early-appearing action patterns into instrumental behaviors.

The development of skills in reaching and manipulation are closely related to the development of such cognitive skills as mental rotation and means-end relationships. When manipulating objects, the subject need to imagine the goal state of the manipulation and the procedures of how to get there. Von Hofsten & Örnkloo [vHO05] studied how infants develop their ability to insert blocks into apertures. The task was to insert elongated objects with various cross-sections (circular, square, rectangular, elliptic, and triangular) into apertures in which they fitted snugly. All objects had the same length and the difficulty was manipulated by using different cross sections. The cylinder fitted into the horizontal aperture as long as its longitudinal axis was vertical, while all the other objects also had to be turned in specific ways. The objects were both presented standing up and lying down. It was found that although infants younger than 18 months understood the task of inserting theblocks into the apertures and tried very hard, they had no idea of how to do it. They did not even raise up elongated blocks, but just put them on the aperture and tried to press them in. The 22-month-old children systematically rose up the horizontally placed objects when transporting them to the aperture and the 26-month-olds turned the objects before arriving at the aperture, in such a way that they approximately fit the aperture. This achievement is the end point of several important developments that includes motor competence, perception of the spatial relationship between the object and the aperture, mental rotation, anticipation of goal states, and an understanding of means-end relationships. These abilities are not independent of each other in a task like this and cannot be totally separated. Motor competence is expressed in actions and actions rely on spatial perception and anticipations of goal states.

The results indicate that a pure feedback strategy does not work for this task. The infants need to have an idea of how to reorient the objects to make them fit. Such an idea can only arise if the infants can mentally rotate the manipulated object into the fitting position. The ability to imagine objects at different positions and in different orientations greatly improves the child’s action capabilities. It enables them to plan actions on objects more efficiently, to relate objects to each other, and plan actions involving more than one object.

7.4.4 Development of Social Abilities

The infant is a social being from birth. Newborns imitate gestures and engage in face-to-face interactions. Such primary intersubjectivity serves to establish strong bonds with caregivers at an age when infants crucially depend on them. From the first months of life, infants understand basic emotions communicated by facial gestures and use such gestures themselves.

During the first year of life, infants become increasingly skilled at understanding the emotions and intentions of other people, and engage in referential communications. Among other things this requires infants to perceive the direction of attention of others. Perceiving what another person is looking at is an important social skill. One can comment on objects and immediately be understood by other people, convey information about them, and communicate emotional attitudes towards them.

Social interaction relies primarily on vision, touch, and proprioception. The mouth, face, eyes, and hands are the primary instruments for such actions. There is an important difference between these action systems and those used for negotiating the physical world. The fact that one's own actions affect the behavior of the person towards whom they are directed creates a much more dynamic situation than when actions are directed towards objects. In addition, anticipating what is going to happen next is less dependent on physical laws as in the object case and more dependent on knowledge of the rules and regularities that govern the other persons actions that in turn is dependent on one's own social behavior and social conventions. In order to master social interaction it is therefore crucially necessary to know the conventions of social interaction and perceive the intentions and emotions of the subject with whom one interacts. Intentions and emotions are readily displayed by elaborate and specific movements, gestures, and sounds which become important to perceive and control. Some of these abilities are already present in newborn infants and reflect their preparedness for social interaction. Neonates are very attracted by people, especially to the sounds, movements, and features of the human face [JM91, Mau85]. They also engage in social interaction and turn-taking that among other things is expressed in their imitation of facial gestures. Finally, they perceive and communicate emotions such as pain, hunger and disgust through their innate display systems [Wol87]. These innate dispositions give social interaction a flying start and open up a window for the learning of the more intricate regularities of human social behavior. Parents show a remarkable talent for responding to the infant's signals and turning them into sophisticated forms of social interaction. [RS99] suggested that this "propensity to express empathy through the echoing of affects and feelings in highly scaffolding ways is part of normal parenting and …the primary source of intersubjectivity".

Important social information is provided by vision. Primarily, it has to do with perceiving the facial gestures of other people. Such gestures convey information about emotions, intentions, and direction of attention. Perceiving what another person is looking at is an important social skill. It facilitates referential communication. One can comment on objects and immediately be understood by other people, convey information about them, and communicate emotional attitudes towards them [CM98, DHM97, MAB97, ST01]. The ability to perceive the gaze direction of others is thus a key component in social communication [MMR98].

Most researchers agree that infants reliably follow gaze from 10-12 months of age [SB75, MRD95, CM98, DFP00, MMD00a, MMD00b, WG02]. A common method has been to determine the side toward which the infants first turn their gaze (see *e.g.* [CM98, MMR98]. Moore et al. [MAB97], for instance, found that some 9-month-olds, and presumably those with more advanced gaze-following skills, will turn in the direction indicated by a live but static face (left or right). Even 3-6-month-old infants have been found to be above chance in following a turning gaze to the correct side [DHM97]. A reasonable conclusion from these studies is that social directional cues can be utilized before 12 months as weak evidence of the probability that some interesting target will be seen to the left or right of the infant. Von Hofsten, Dahlström, & Fredriksson [vHDF05] used an eye tracker (TOBII) to study 12-month-old infants' ability to perceive gaze direction in static video images. The images showed a woman who performed attention directing actions by looking and/or pointing towards one of 4 objects positioned in front of her (2 on each side). They found that the infants clearly discriminated the gaze directions to the objects located 10° apart, on the same side of the model. The infants spent more time looking at the attended objects than the unattended ones and they shifted gaze more often from the face of the model to the attended object than to the un-attended objects. In all conditions the infants spent most of the time looking at the model's face. This tendency was especially noticeable in the pointing-only condition and the condition where the model just looked straight ahead.

Humans possess a unique ability to underline their direction of attention by pointing [Tom06]. It is performed with different goals depending on the context. Bates [Bat75] discussed pointing as a way to to share the attention in an object (declarative) or to request something (imperative) [Bat75]. The difference is that during imperative pointing infants want to get an object, thereby making the pointing an instrumental gesture, while during declarative pointing they want to share the attention of an interesting object with another person using a socially communicative gesture. This distinction becomes important if we consider that neither of the two types of pointing has been seen naturally in animals, but the imperative pointing can be learned by apes in captivity [Tom06; PBG03] and in children with autism. No declarative pointing has been observed in these groups. Still some studies show that the main function of pointing is the declarative one, which cast doubt on the hypothesis that pointing develops from grasping as some authors thought [Vyg78]. Pointing has also other functions apart of sharing attention or requesting an object. Infants as young as 12 months can use it to provide information to adults (the location of an object that the adult was looking for) [LCS+06] and there are more possibilities, like asking for information about objects (such as names), indicating a direction, creating imaginary shapes or even to show inferred referents (i.e. pointing to an empty chair to refer to the person who usually seats there) [Kit03]. With this in mind it is not strange that pointing has been studied extensively because of its connection with language. Some studies show that pointing at 12 months predicts speech production rates at 24 months [CCL+91] and that the combination of pointing and a word which differs from the object signed precedes two-word sentences, the first grammatical construction [GMB03]. Also, some researchers indicate how index finger extension is correlated with production of syllabic sounds [Mas03] and that pointing can be the first way to individuate the visual object with a sound [But03].

As we get more information about pointing, we are left with many unanswered questions. We know that infants start pointing between 8 and 13 months [But03] but there is a current discussion on why infants start to point. The onset may be innate or start through imitation when infants see others pointing. The onset can also be conditioned by the presence of the object they want (imperative pointing) or by the parent's positive reaction and shared attention (declarative pointing) [CNT98]. There is also an active discussion whether the declarative or imperative pointing comes first. When it comes to understand others' pointing, some authors think that infants probably comprehend pointing one month before they perform pointing themselves [But03] and others state that infants start pointing before they follow other's pointing [Mat03].

The most important perception-action systems that serve social interaction is speech. Like other action systems, speech has both a perceptual and a productive side. Perception of certain aspects of speech, seem to occur in the womb already, and newborn infants have been shown to prefer their mother's voice [DF80]. Because of the lowpass filtering of the human voice in the womb, it is presumably the prosody of the voice rather than any other more detailed property that neonates recognize. There is good evidence that infants are sensitive to prosodic structure and that this sensitivity is present in the newborn [Jus92]. Also the phonemic structure develops early. By 4 month of age, infants seem to be able to distinguish between virtually any pair of stimuli that crosses phoneme boundaries [Kuh94].

The research on early development of speech shows that the productive capabilities of speech clearly lag the perceptual ones (see *e.g.* [Men83]). Thus, human infants can perceive speech before the can speak or babble. On the other hand, phylogeny has prepared the human child for the task of speaking. The morphology of the human vocal tract has been altered relative to that of other primates in a way that facilitates speech [CLM95]. Babbling is, furthermore, dominated by the cyclical opening and closing of the mandible in a way that is also characteristic of sucking. MacNeilage & Davis [MD93] argued that many of the articulatory regularities in the sound patterns of babbling and early speech can be attributed to properties of this mandibular cycle. During the second half of life infants spend much of their time awake exercising babbling sounds. They also discover the communicative value of speech sounds and use them in their social interactions much before they can articulate specific words. In addition to this, infants start pointing at around 11 months of age, "a crucial step on the road to language" [But??]. Pointing often starts when objects are named, an example of that language and planned directed actions are connected.

8 The Co-Development of Morphology and Cognition

9 Summary

We conclude by attempting a short summary of all of the many issues addressed in this part of the deliverable. In doing so, we will highlight the key points and, where relevant, provide the timeline for development of certain abilities.

9.1 Action as the Organizing Principle in Cognitive Behaviour

The ultimate function of cognition is to guide actions. Movements of neonates are based on actions, not reactions or reflexes. Actions are specifically-motivated and are guided by prospective information. Actions are defined by the goals (or goal state) not by the movements or trajectories that form them. Perception provides direct information on what is going to happen next (not just what is happening right now) and cognition extends perception and provides a longer timeframe in predicting what is going to happen next.

Perception and action are mutually dependent. Actions require perception to guide them but actions also form part of the perceptual process. All actions have perceptual functions. The perceptual system, the effector system (the system morphology), and the neural system are tailored to solve specific action problems.

9.2 Phylogeny: Core Abilities and Core Knowledge

Development depends on the presence of some built-in (innate) abilities provided by phylogeny. These innate abilities present themselves through morphological pre-structuring, pre-structuring of the motor system, pre-structuring of the pereceptual system, sensory-motor couplings, and innate (or core) perceptual and conceptual abilities.

Although morphology, the motor system, and the perceptual system are all pre-structured, this does not mean they are fixed. Quite the opposite: changes in morphology can produce adjustments to the perceptual system to improve the extraction of information to control specific actions. Initially, activation of muscles is organized into functional synergies which reduce the number of effective degrees of freedom. For example, arm extension and finger extension are linked in neonates but is later decoupled during the development stage. The vision system is structured pre-natally in two stages of neural mapping, both of which are necessary. The first stage is determined by the genotype and the second by the activity of the foetus. In the first stage, neural pathways between the retinal and LGN & superior colliculus develop that preserve in a very crude way the retinal topographies. In the second stage, structured activities at the retinal level refine this mapping through competitive or Hebbian interactions. Spontaneous movement of the foetus may provide these structured stimuli (*e.g.* moving hands across the eyes produces shadows that may both assist in mapping the visual system and the coupling of the visual and motor systems).

An important part of core knowledge is concerned with people. Infants are attracted by other people, they have the ability to recognize them and their expressions, communicate with them, and perceive the goal of an action. Infants can perceive and understand the relationship between themselves and the spatial layout of their surroundings and their orientation relative to external landmarks. Infants can understand small numerical quantities and larger numbers in an approximate manner.

Core knowledge systems are limited in a number of ways: they are domain specific, task specific, and encapsulated. Core knowledge systems serve as the foundation for later development but they also continue to operate effectively in tandem with abilities developed later. Core abilities are not fixed and rigid but develop and refine with time.

The following is an non-exhaustive list of the innate core abilities of the neonate.

Objects

- Distinguish between relative and common motion in the visual field (based on either object or observer movement).
- Group motions that share common motion (hence distinguish between occluding and partially occluded objects).
- Ascribe 'objecthood' to part of the visual field that have persistent and well-defined outer boundaries.
- Track objects through occlusion by extrapolation of motion, only saccading after the expected transition time behind the occluder.

Numbers

- Modality-independent ability to disguish between one, two, and three entities.
- Modality-independent ability to disguish between groups of entities exhibiting relatively gross differences in number.

Space

- Navigate based on dynamic egocentric representations based on information of low complexity and number.
- Reorient based on local view-dependent landmarks (rather than global scene representations).

People

- Attend to sounds, movements, and features of the human face.
- Detect mutual gaze (*i.e.* eye-contact).

9.3 Development

Interaction is crucial for development: perception, cognition, and motivation develop at the interface between neural processes and actions, are a function of both (and, hence, are a function of the morphology of the system). Consequently, bio-mechanical constraints affect and are affected by the nervous system and the way actions are performed.

In the first few months of post-natal life, there is a massive increase in the connectivity in the cerebral cortex and the cerebellum, after which a process of self-organization produces new forms of perception, action, and cognition. The emergence of new forms of action always relies of several other contributing developments.

Development depends crucially on motivations which define the goals of actions. The two most important motives that drive actions and development are social and explorative. Social motives include comfort, security, and satisfaction. There are at least two exploratory motives, one involving the discovery of novelty and regularities in the world (infant interest declines rapidly with repeated exposure to new objects), and one involving finding out about the potential of one's own actions.

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

Repetitive practice of new actions is not focused on establishing fixed patterns of movement but on establishing the possibilities for prospective control in the context of these actions.

9.4 Development of Perception

There are three types of perceptual development: the spontaneous learning that facilitates the detection of structure (regularity) in sensory flow, the process of selecting the information relevant for guiding actions, and the interpretation of percepts on the basis of the action capabilities.

The following shows the timeline for the development of visual processing in the neonate.

M0 Visual acuity is only 3-5% of that of an adult
M0 Object-directed action
M5 Visual acuity reaches adult-like levels
M1 Ability to process colour
M2 Ability to achieve ocular convergence for objects beyond 20 cm
M2 Ability to process motion information
M3 Sensitivity to motion parallax
M3 Ability to perceive binocular depth

9.5 Development of Posture and Locomotion

Effecting and maintaining stable orientation is a pre-requisite for other purposeful actions: posture control is a limiting factor in motor development. Gravity provides a frame of reference and the otolith sensors in humans provide information on the direction of gravity. Vision and proprioception also provide important information for stability. Vision is crucial for supporting balance prospectively; control of posture must also factor in one's own movements prospectively. It is used principally for detecting small body displacements.

The following shows the timeline for the development of posture control.

M3 First sign of being able to control gravity (lift and control of head when lying prone)
M6-8 Sitting (consistent control of sway of head and trunk);
Transfer from two-handed reaching to one-handed reaching
M12 Infants who can stand are very sensitive to peripheral visual information
in controlling body displacement and balance; sensitivity improves with experience

9.6 Development of Looking

Development of ocular motor control is one of the earliest skills to appear and is crucial for gaze, attention, and social communication.

Gaze control involve both head and eye movement, and uses visual, vestibular, and proprioceptive information. Gaze involves (a) high speed saccadic eye movements to direct the eyes and shift gaze to significant visual targets, and (b) smooth pursuit eye movements to stabilize the target. Smooth pursuit requires anticipation (predictive control).

Compensatory gaze adjustment unrelated to fixation is also required: (a) visual to minimize retinal slip (effective on slow optical change up to approx. 0.6 Hz) and (b) vestibular to stabilize gaze in space (operates best above 1 Hz).

The following shows the timeline for the development of looking in the neonate.

M0 Vestibular gaze stabilization (to compensate for head movement)
Saccadic eye movements, ability to engage and disengage attention; develops rapidly in M1-6
Very limited smooth pursuit ability
Attentional processes are also present: gaze directed toward attractive objects
(novel appearance or events);
Systematic scanning of surroundings doesn’t emerge until after M24 (pre-school age)
M1½ Rapid improvement of smooth pursuit
M3-4 Infants achieve adult level of smooth pursuit

9.7 Development of Reaching and Manipulation

Early reaches are segmented, with distinct acceleration and deceleration phases, unlike adult reaching which exhibits a single bell-curve velocity profile for the entire reach-grasp action.

The first grasps are palmar and use the whole hand; differentiated finger grasping develops after months 9-10.

Development of skills in reaching and manipulation are closely related to cognitive skills such as mental rotation and means-ends relationships.

The following shows the timeline for the development of reaching and grasping.

M0 Visual control of arm but no control of fingers for grasping
Arm and finger motions governed by global extension and flexion synergies
M2 Coupling of global arm and finger motions is broken: hand is fisted when extending the arm
M2-3 Open hand when reaching, but only when visually guided; hand closing when close to object
M4-5 Reaching and grasping
M5 Hand not adjusted to size of object when reaching
M9 Onset of adjustment of hand size when reaching; hand closed when in vicinity of object
M9-10 Differentiated finger grasping, *e.g.* pincer grasp
M13 grasping starts when reaching: *i.e.* one integrated reach-grasp act

9.8 Development of Social Abilities

New-borns are attracted by people, especially sound, movements, and features of human faces, and they imitate gestures and engage in face-to-face communication. They understand emotional states and facial gestures, and use such gestures to communicate. In this, the need to perceive the direction of gaze and attention of others is very important. Since one’s own actions directly impact on the behaviour of others to whom they are directed, a much more dynamic situation than when manipulating objects is created.

Prospection requires an understanding of the rules of interaction (not just physical laws and regularities) which include one’s own behaviour. This highlights the need to be able to perceive the emotional state and intention of others.

The following shows the timeline for the development of the perception of gaze in the neonate.

M3-6	Can perceive the correct hemisphere of gaze
M10-12	Can follow gaze

Part III
Neurophysiological and Psychological Models

10 Modalities and Models of Perception

Since the early Eighties, the dominant view on the cortical processing of visual information has been the ‘what’ and ‘where’ theory, as formalized by Ungerleider and Mishkin [UM82]. According to these authors, the ventral stream has its main role in object recognition, while the dorsal stream analyzes object’s spatial location. This point of view was in accordance with the classical notion of the parietal cortex as the site for unitary space perception, used for all purposes: for walking, for reaching objects, for describing a scene verbally. Lesions of this lobe and, in particular, of the inferior parietal lobule produce a series of spatial deficit ranging from space distortions to spatial neglect.

Since 1991, Milner and Goodale have argued against this theory, emphasizing the pragmatic role of the dorsal stream. This point of view, primarily triggered by clinical data, has been subsequently substantiated by neurophysiological evidence. The posterior parietal cortex, as pointed out also by Milner and Goodale [MG95] is constituted by a mosaic of independent areas. If one of these areas were the hypothetical space master center, it should be also the center of a series of convergent and divergent connections. It should receive inputs from the occipital lobe and distribute its output to a variety of other brain centers: oculomotor centers for looking at the objects, areas controlling walking for navigating in the environment, and so on. The evidence is exactly the opposite. The connections of parietal lobe with the frontal lobe as well as with subcortical centers are remarkably segregated ([AAEM90, CGR89, MCGR86, PP84]). For example, the connections of parietal area LIP (lateral intraparietal) are exclusively or almost exclusively with Brodmann area 8 (frontal eye fields, FEF). Both these areas are related to oculomotion. Area LIP, in contrast, does not send any input to areas related to arm movements. Thus there is no evidence of a unique supramodal “space area” within the posterior parietal cortex. Space perception appears to derive from the joint activity of a series of sensorimotor fronto-parietal circuits, each of which, according to its own motor purposes, encodes the spatial location of an object and transforms it into a potential action (see [RFG97, RFFG97]).

The idea of a motor role for the posterior parietal cortex is by no means new. Since the pioneering studies carried out by Hyvarinen, Mountcastle and their co-workers [HP74, MLG
75] it is well known that different sectors of the posterior parietal cortex are involved in the control of arm, hand and eye movements. However, the ‘motor’ role was somehow underestimated in light of a purely ‘spatial’ characterization of the visual information reaching these sectors of the parietal cortex.

Milner and Goodale [MG95] make two major points: 1) The dorsal stream processes visual information for motor purposes; 2) Action and perception are two completely separate domains, the latter being an exclusive property of the ventral stream. While a consistent set of neurophysiological data confirm the ‘pragmatic’ role of the visual information processed in the dorsal stream, and thus corroborates the theoretical views of Milner and Goodale [MG95], a series of neurophysiological, neuropsychological and brain imaging studies suggests that the dichotomy proposed by Milner and Goodale between action and perception is probably too rigid.

Among the arguments in favor of the ‘pragmatic’ role of the visual information processed in the dorsal stream, are the functional properties of the parieto-frontal circuits. For reason of space we will review here in some detail only the functional properties of two circuits, that formed by area LIP and FEF, and that constituted of parietal area VIP (ventral intraparietal) and frontal area F4 (ventral premotor cortex). The same functional principle is valid, however, also for the other circuits.

The LIP-FEF circuit contains three main classes of neurons: neurons responding to visual stimuli (visual neurons), neurons firing in association with eye movements (motor neurons), and neurons with both visual- and movement-related activity (visuomotor neurons) [AG89, BG85, GS89]. Neurons responsive to visual stimuli respond vigorously to stationary light stimuli. Their receptive fields (RFs) are usually large. Movement-related neurons fire in relation to ocular saccades, most of them discharging before the saccade onset. Visuomotor neurons have both visual- and saccade-related activity. Visual RFs and 'motor' fields are in register, that is, the visual RF corresponds to the end-point of the effective saccade. Visual responses in both LIP and FEF neurons are coded in retinotopic coordinates [AG89, GS89]. In other words, their RFs have a specific position on the retina in reference to the fovea. When the eyes move, the RF also moves. Most LIP neurons have, however, an important property. The intensity of their discharge is modulated by the position of the eye in the orbit (orbital effect). Now, if the position of the RF on the retina and the position of the eye in the orbit are both known, one can reconstruct the position of the stimulus in spatial (craniocentric) coordinates. Thus, although the firing of a neuron does not specify by itself the position of the triggering stimulus in space, the spatial location of stimulus can be derived from the discharge intensity of different neurons [BASG95].

As in the LIP-FEF circuit, neurons in the VIP-F4 circuit can be subdivided into three main classes: sensory neurons, movement-related neurons, and sensorimotor neurons. The majority of them belong to the last category. Movement-related neurons and sensorimotor neurons are activated by head movements, face movements, or arm movements. Sensory and sensorimotor neurons respond to tactile or to tactile and visual stimuli. The visual RFs of these neurons are anchored to the tactile ones regardless of eye position [GSPR83, GFL+88]. F4 neurons fire tonically at the presentation of stationary three-dimensional objects within monkey peripersonal space. A very intriguing finding is that some of these tonically discharging neurons continue to fire when, unknown to the monkey, the stimulus previously
presented has been withdrawn, and the monkey 'believes' that it is still near its body. Space representation in the premotor cortex can be generated, therefore, not only as a consequence of an external stimulation but also internally on the basis of previous experience [GHG97].

If we now compare the properties of the VIP-F4 circuit with those of the LIP-FEF circuit, we find a common aspect and some important differences. The common aspect is that coding of space is not devoted to a multiplicity of purposes but is specifically directed to a particular motor goal: eye movements in the case of the LIP-FEF circuit, body-part movements in the case of the VIP-F4 circuit. The different aspect is the way in which spatial information is obtained. For eye movements, space is coded by retinotopic neurons which change their activity with the position of the eyes in the orbit. For head, arm, and hand movements, space is coded in body-centered coordinates (neurons signal the location of a stimulus with respect to a specific body-part). The difference between the properties of the LIP-FEF circuit and those of the VIP-F4 circuit is probably a cue for understanding why there is no multipurpose space map. The various motor effectors need different information and have different sensory requests. These cannot be provided by a unique map. Furthermore, the sensorimotor transformations necessary for organizing different types of movements must obviously have appeared in evolution before conscious space perception. Thus, conscious space perception derived from a conjoint action of the pre-existent spatial maps, rather than from the appearance of a new multipurpose map. The appearance of a new map specific for conscious space perception would entail an enormous rewiring and a complete reorganization of the whole cerebral cortex. Evolutionary speaking, such a rearrangement is extremely unlikely.

Summing up, within the dorsal stream, there are parallel cortico-cortical circuits, each of which elaborates a specific type of visual information in order to guide different types of action. The peculiarity of these circuits resides in the fact that different effectors are provided with the most suitable type of visual information required by their motor repertoire. This firm connection between vision and action seems to be the organizing principle within the circuitry connecting the parietal with the agranular frontal cortex of the monkey.

11 Perception / Action Dependencies

A strong point made by Milner and Goodale [MG95] maintains that in the primates' visual system there is a sharp distinction between the role played by the dorsal and the ventral stream of visual processing: the dorsal stream would be mainly involved in the on-line control of actions, while the ventral stream would be the exclusive source of information for perception and semantics. Several lines of evidence seem to point to an important involvement of the motor system in supporting processes traditionally considered to be 'high level' or cognitive, such as action understanding, mental imagery of actions, perceiving and discriminating objects. A first example is provided by the discovery of a population of neurons in the monkey ventral premotor cortex (mirror neurons) that discharge both when the monkey performs a grasping action and when it observes the same action performed by other individuals [GFFR96]. Mirror neurons would provide the neurophysiological basis for the capacity of primates to recognize different actions made by other individuals: the same motor pattern which characterizes the observed action is evoked in the observer and activates its own motor repertoire. This matching mechanism, which can be framed within the motor theories of perception, offers the great advantage of using a repertoire of coded actions in two ways at the same time: at the output side to act, and at the input side, to analyse the visual percept. This matching system has also been demonstrated in humans. Transcranial Magnetic Stimulation (TMS) of the motor cortex of subjects observing hand actions made by the experimenter determined an enhancement of motor evoked potentials (MEPs) in the same muscular groups that were used by the experimenter in executing those actions [FFPR95]. This means that when we observe an action we utilize, as monkeys do, the repertoire of motor representations used to produce the same action. Another example of the involvement of the dorsal stream in cognitive functions is motor imagery. Imagining a grasping action is a cognitive task that requires a conscious, detailed representation of the movement. Several PET studies have shown that during motor imagery of grasping actions premotor and inferior parietal areas are strongly activated [DPJ+94, GAFR96]. Furthermore, Parsons et al. [PFD+95] demonstrated in a PET study that motor imagery used for visual hand shape discrimination activates premotor and posterior parietal cortex. Further evidence supporting the notion of the involvement of the dorsal stream in cognitive tasks is provided by an elegant neuropsychological study by Sirigu et al. [SDC+96]. Patients with lesions restricted to the posterior parietal cortex were selectively impaired at predicting through mental imagery the time necessary to perform differentiated finger movements. The role played by handedness in performing cognitive tasks is another example of the involvement of motor processes in perceptual functions[dSS97] showed that right- and left-handed normal subjects used an internal simulation of the movement of their dominant hand in order to discriminate between observed screwing and unscrewing screwdrivers. In another series of experiments [GDG98a, GDG98b], normal subjects were required to judge handedness of pictures of hands and fingers assuming different postures. The results showed that the presentation of postures that hand and fingers commonly assume at rest, or when interacting with objects, facilitated the responses with respect to the presentation of less usual hand-finger postures, even when the latter were richer in visual cues useful for handedness recognition. Once again procedural motor knowledge was employed to solve a cognitive task. Taken together, all these results seem to contradict a sharp distinction between an 'acting brain' and a 'knowing brain'. Among the processes traditionally considered to be 'high level' or cognitive, selective attention is one of the most important. It refers to the capability of selecting a particular stimulus according to its physical properties, way of presentation, or previous contingencies and instructions. After selection, the stimulus is processed and, if convenient for the individual, acted on. According to the scenario for space representation described above, a problem is how the different sectors of space representation can increase their efficiency in processing visual stimuli in order to select some of them and discard others.

The traditional view is that selective attention is controlled by a supramodal system 'anatomically separate from the data processing systems' ([PP90], p. 26). Like the sensory and motor systems, this 'attention system' performs operations on specific inputs. It interacts with other centers of the brain but maintains its own identity [PP90]. On the basis of data obtained from brain imaging experiments

[CMD+90, CMD+91, PPFR88], it has been suggested that the attention system is not unitary but consists of at least two independent systems: a posterior one subserving spatial attention and an anterior one devoted to attention recruitment and control of brain areas involved in complex cognitive tasks [PD94].

Another view of selective attention is that it derives from mechanisms that are intrinsic to the circuits underlying perception and action. Attention is modular, and there is no need to postulate control mechanisms anatomically separate from the sensorimotor circuits. This account of selective attention was originally formulated for spatial attention (premotor theory of attention; [RC87, RRDC87] and it is deeply rooted in the idea that space is coded in a series of parieto-frontal circuits working in parallel and that the coordinate frame in which space is coded depends on the motor requirements of the effectors that a given circuit controls (see [RRS94]). Given this strict link between space coding and action programming, the premotor theory of attention postulates that spatial attention is a consequence of an activation of those cortical circuits and subcortical centers that are involved in the transformation of spatial information into action. Its main assumption is that the motor programs for acting in space, once prepared, are not immediately executed. The condition in which action is ready but its execution is delayed corresponds to what is introspectively called spatial attention. In this condition, two events occur: (a) There is an increase in motor readiness to act in the direction of the space region toward which a motor program was prepared, and (b) the processing of stimuli coming from that same space sector is facilitated. There is no need, therefore, to postulate an independent control system. Attention derives from the mechanisms that generate action. Although, in principle, all circuits responsible for spatially directed action can influence spatial attention, there is no doubt that in humans the central role in spatial attention is played by the circuits that code space for programming eye movements. Experiments in which the relations between attention and eye movements were either indirectly or directly tested showed that the two mechanisms interact: Any time attention is directed to a target, an oculomotor program toward that target is prepared. Particularly significant in this respect are experiments in which the relations between attention and eye movements were directly tested [SRCR95b, SRCR95a]. Sheliga and coworkers instructed normal participants to pay attention to a given spatial location and to perform a predetermined vertical or horizontal ocular saccade at the presentation of the imperative stimulus. Results showed that the trajectory of ocular saccades in response to visual or acoustic imperative stimuli deviates according to the location of attention. The deviation increased as the attentional task became more difficult. Note that if spatial attention were independent of oculomotor programming, ocular saccades should not be influenced by location of attention. In a recent experiment, the role of oculomotion in orienting of attention was investigated by dissociating perceptual from motor capabilities [CNF04]. If a causal relationship links oculomotion and orienting of attention, any constraint limiting eye movements should abolish, or at least reduce, attentional benefits in the region of the spatial field barely reachable by the eye. On the contrary, if attention is a purely cognitive process, then no effects are expected to arise from oculomotor constraints. Subjects were submitted to a spatial attention orienting task, performing it in monocular vision and having the head rotated in such a way that the eye was kept at an extreme position in the orbit. This position limited the execution of a saccade toward the temporal hemifield, whereas it allowed saccadic execution toward the nasal hemifield. Results showed that orienting of attention was normal in the nasal but not in the temporal hemifield, indicating that eyes and attention show a common limit stop.

Whereas in primates eye movements are certainly the most important mechanism for selecting stimuli, there are also circumstances (*e.g.*, stimuli presented very close to the face) in which eye movements are not crucial in selecting stimuli in space. In these circumstances, spatial attention should depend on circuits other than those related to eye movements. Probably the best documented evidence in favor of spatial attention not related to eye movements is that deriving from experiments conducted by Tipper et al. [TLB92]. They studied, in normal participants, the effect of an irrelevant stimulus located in or out of the arm trajectory necessary to execute a pointing response. The results showed that an interference effect was present only when the distractor was located in the trajectory of the arm. Control experiments suggested that the effect was not due to a purely visual representation of the stimuli or to spatial attention related to eye movements. Rather, the organization of the arm-hand movement determined a change in the attentional relevance of stimuli close to the hand or far from it. In the frame of premotor theory of attention, Craighero and colleagues [CFRU99] assumed that allocation of attention to a graspable object is a consequence of preparing a grasping movement to that same object. The authors predicted that, when a specific grasping movement was activated, there would be both an increase in the motor readiness to execute that movement and a facilitation in visual processing of graspable objects the intrinsic properties of which are congruent with the prepared grasping. In the experiment normal subjects were required to grasp a bar after the presentation of a visual stimulus whose orientation was either congruent or incongruent with that of the bar. The results supported the hypothesis. The detection of a visual object was facilitated by the preparation of a grasping movement congruent with the object's intrinsic properties. This finding strongly suggests that the premotor theory of attention is not limited to orienting attention to a spatial location but can be generalized to the orienting of attention to any object that can be acted on.

12 Summary

Conventional thinking has it that visual information is processed for object recognition in the ventral stream and for spatial location (to be used in motor control) in the dorsal stream [UM82], and that the posterior parietal cortex acts as a unique site for space perception.

However, recent evidence suggest that, on the contrary, space perception is not the result of a single circuit, and in fact derives from the joint activity of several fronto-parietal circuits, each of which encodes the spatial location and transforms it into a potential action in a distinct and motor-specific manner [RFG97, RFFG97]. In other words, *the brain encodes space not in a single unified manner — there is no general purpose space map — but in many different ways, each of which is specifically concerned with a particular motor goal.* Different motor effectors need different sensory input: derived in different ways and differently encoded in ways that are particular to the different effectors. Conscious space perception emerges from these different pre-existing spatial maps.

As an example of these distinct space perception / movement mechanisms, consider the Lateral Intraparietal (LIP) area and the Brodmann area 8 Frontal Eye Fields (FEF). The LIP-FEF circuit contains mainly visual neurons, motor neurons, and visuo-motor neurons. While the receptive fields of both the visual and motor neurons (for saccade movements) are effectively registered, in that they are both defined in a retinocentric frame of reference, the location of a stimulus *in a craniocentric frame of refence* can still be inferred because the intensity of discharge of the visual neurons is modulated by the position of the eye in its orbit (and, hence, modulated by the saccade motor neural activity). Furthermore, not only is spatial information derived and encoded in action-specific mechanisms, there is also evidence that the distinction (or disjointedness) between perception for action control and perception for semanic understanding is not valid. In fact, it appears that the motor system is very much involved in the semantic understanding of percepts.

For example, there is the recent discovery of the so-called mirror neurons in the ventral premotor cortex that discharge both when, for instance, a grasping action is performed and when the same action is observed being performed by others [GFFR96]. In addition, the cognitive operations of imagining (or visualizing) a grasping action [DPJ+94, GAFR96] or discriminating between hand shapes [PFD+95] also involves the dorsal stream and the premotor and inferior parital areas.

Finally, selective attention too it seems is not a unitary system but rather a process that derives from the several cortical circuits and subcortical centres that are involved in the perception-action dependent transformation of spatial information into movements or actions. Thus, attention derives from the mechanisms that generate action: it is the simultaneous occurence of a readiness to act in some spatial region and a predisposition to process stimuli coming from that region. Thus, spatial attention is yet another example of the co-dependency of perception and action.

For example, spatial attention is dependent on oculomotor programming: when the eye is positioned close to the limit of its rotation, and therefore cannot saccade in any further in one direction, visual attention in that direction is attenuated [CNF04].

This premotor theory of attention applies not only to spatial attention but also to selective attention in which some object rather than others are more apparent. For example, the ability to detect an object is enhanced when features or the appearance of the object coincide with the grasp configuration of a subject preparing to grasp an object [CFRU99]. In other words, the subject’s actions conditions its perceptions.

Part IV
Work-in-Progress Models of Cognition

[This part will summarize each of the partners' individual computational models of cognitive skills. The first section just sets the scene. This part has not yet been completed.]

13 The Space of Ontogeny: Action & Prospection

Since development is a temporally-extended event, it allows us to study cognition in an incremental way without having to understand the complete scheme *ab initio*. It also gives us a way to choose an early point of departure in the endeavour to understand cognition and then to make progress as the system itself develops. The legitimacy of this methodology is supported by experience in studying newborn infants[Hut90, Atk00, vHR97] as detailed above.

The question is, of course, how are we to do this? In RobotCub we view the process of development as a sort of traversal of a two-dimensional space of ontogeny, one dimension corresponding to prospection (or the degree of prospective control required to develop and accomplish a skill), and the other dimension corresponding to the degree of sophistication of the actions that must be recruited by the system to develop these (increasingly cognitive) skills.

Thus, we begin with the immediate time scale (*e.g.* motor control, sensory mapping, *etc.*), aspects of prospective control (*e.g.* reaching/grasping moving objects, tracking/eye movement, anticipation), followed by the more elaborate predictions required for manipulating objects (*e.g.* grasping according to shape and use), and finally, towards skills requiring deliberation and prediction such as communication, imitation, and complex manipulation involving tools.

The development of prospective control includes the following.

- Discovering the manipulation abilities of its own body, learning how to crawl, to bend, to reach for static and moving targets, and to balance when manipulating objects while crawling or sitting.

- Discovering and representing the shape of objects, learning to recognize and track visually static and moving targets, and discovering and representing object affordances (*e.g.* the use of tools).

- Recognizing manipulation abilities of others and relating those to one's own manipulation abilities, learning to interpret and predict the gestures of others, learning new motor skills and new object affordances by imitating manipulation tasks performed by others, and learning what to imitate and when to imitate others' gestures.

- Learning to regulate interaction dynamics, including approach, avoidance, turn-taking, and social spaces, and learning to use gesture as a means of communication.

- Developing 'personalities' *via* autobiographic memory based on interaction histories, learning about meaningful events in the lifetime of the robot and sharing memory (events) during interaction.

The actions we intend to recruit (and develop) include the following:

- Locomotion
- Eye-head-hand coordination

- Bimanual cooperation
- Affordance
- Imitation
- Gestural communication

14 Types of Action

14.1 Locomotion

14.1.1 Crawling

14.1.2 Sitting

14.2 Eye-Head-Hand Coordination

14.2.1 Oculomotor Gaze Control

14.2.2 Visually-guided Reaching and Grasping

[This text is just an example of the type of material we intend to include. Ideally, we will also add in a summary of the technical details, complete with the relevant mathematical exposition of the theory.]

We have already remarked on the co-dependency of perception and action in biological systems. Perceptual development is determined by the action capabilities of a developing child and what observed objects and events afford in the context of those actions [vH04, GP00]. It is worth reinforcing this again, especially in the light of recent neurological evidence. For example, the presence of a set of neurons — mirror neurons — is often cited as evidence of the tight relationship between perception and action [GFFR96, RFGF96]. Mirror neurons are activated both when an action is performed and when the same or similar action is observed being performed by another agent. These neurons are specific to the goal of the action and not the mechanics of carrying it out [vH04].

In summary, the development of action and perception, the development of the nervous system, and the development (growth) of the body, all mutually influence each other as increasingly-sophisticated and increasingly prospective (future-oriented) capabilities in solving action problems are learned [vH04].

An example of a system which exploits this co-dependency in a developmental setting can be found in [MSK99]. 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. The system operates as follows. By assuming that a fixation point represents the object to be reached for, the reaching is effected by mapping the eye-head proprioceptive data to the arm control parameters. The control itself is implemented as a multi-joint synergy by using the control parameters to modulate a linear combination of basis torque fields, each torque field describing the torque to be applied to an actuator or group of actuators to achieve some distinct equilibrium point where the acuator position is stable. That is, the eye-hand motor commands which direct the gaze towards a fixation point are used to control the arm motors, effecting what is referred to in the paper as “motor-motor coordination”. The mapping between eye-head proprioceptive data (joint angular positions) and the arm control parameters is learned by fixating on the robot hand during a training phase.

14.3 Bi-manual Coordination

14.4 Affordances

14.4.1 Learning Affordances by Exploration and Experiment

[This text is just an example of the type of material we intend to include. Ideally, we will also add in a summary of the technical details, complete with the relevant mathematical exposition of the theory.]

This section describes a biologically-motivated system, modelled on brain function and cortical pathways and exploiting optical flow as its primary visual stimulus, which demonstrates the development of object segmentation, recognition, and localization capabilities without any prior knowledge of visual appearance though exploratory reaching and simple manipulation [MF03]. The system also exhibits the ability to learn a simple object affordance and use it to mimic the actions of another (human) agent.

The working hypothesis is that action is required for object recognition in cases where the system has to develop the object classes or categories autonomously. The inherent ambiguity in visual perception can be resolved by acting upon the environment that is perceived. Development starts with reaching, and proceeds through grasping, and ultimately to object recognition.

Training the arm-gaze controller is effected in much the same way as in [MSK99] but in this case, rather than using colour segmentation, the arm is segmented by seeking optical flow that is correlated with arm movements (specifically, during training, by correlating discontinuities in arm movement as it changes direction of motion with temporal discontinuities in the flow field.

Segmentation of (movable) objects is effected also by optical flow by poking the object and detecting regions in the flow field that are also correlated with arm motion, but which can’t be attributed to the arm itself. Objects that are segmented by poking can them be classified using colour histograms of the segmented regions.

A simple affordance — rolling behaviour when poked — is learned by computing the probability of a normalized direction of motion when the object is poked (normalization is effected by taking the difference between the principal axis of the object and the angle of motion).

The effect of different poking gestures on objects is then learned for each gesture by computing the probability density function (a histogram, in effect) of the direction of motions averaged over all objects. There are four gestures in all: pull in, push away, backslap, and side tap.

When operating in a non-exploratory mode, object recognition is effected by colour histogram matching, localization by histogram back-projection, and orientation by estimating the principal axis by comparison of the segmented object with learned prototypes.

The robot then selects an action (one of the four gestures) by finding the preferred rolling direction (from its learned affordances) adding it to the current orientation and then choosing the gesture which has the highest probability associated with resultant direction.

Mimicry (which differs from imitation, the latter being associated with learning new behaviour, and the former with repeating known behaviour [Bil02]) is effected by presenting the robot with an object and performing an action on it. This “action to be imitated” activity is flagged by detecting motion in the neighbourhood of the fixation point, reaching by the robot is then inhibited, and the effect of the action of the object is observed using optical flow and template matching. When the object is presented again a second time, the poking action that is most likely to reproduce the rolling affordance is selected. It is assumed that this is exactly what one would expect of a mirror-neuron type of representation of perception and action. Mirror neurons can be thought of as an "associative map that links together the observation of a manipulative action performed by someone else with the neural representation of one's own actions".

14.4.2 Learning Affordances by Imitation

14.5 Imitation

14.5.1 Goal-directed Pointing and Reaching

14.5.2 Functional Imitation of Arm Motion

14.5.3 Role Reversal in Demonstration and Imitation

14.6 Gestural Communication

14.6.1 Regulation of Interaction Dynamics

Part V
A Research Roadmap

We come finally to the realization of cognition in the iCub . Part V of Deliverable 2.1 brings together everything has has been discussed so far to create a coherent open software system that encapsulates the emergent developmental philosophy of the consortium, and exercises its scientific theories of cognition, while at the same time providing a framework to allow others in the community to incorporate their own theories, exploiting as much or as little of the iCub cognitive system as they want.

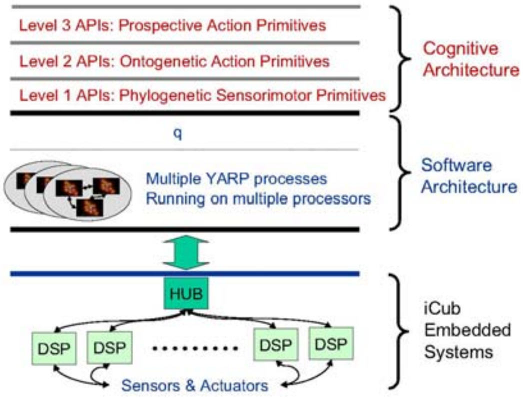

Level 3 APIs: Prospective Action Primitives
Level 2 APIs: Ontogenetic Action Primitives
Level 1 APIs: Phylogenetic Sensorimotor Primitives
Cognitive Architecture
q
Multiple YARP processes
Running on multiple processors
Software Architecture
HUB
DSP
DSP
DSP
DSP
Sensors & Actuators
iCub Embedded Systems

Figure 3: The layers of the iCub architecture.

Note that we are concerned in this deliverable only in the explicitly cognitive aspects of the overall iCub system. These are intrinsically linked to other aspects, specificially the software architecture and the iCub's embedded systems. This relationship is expressed schematically in Figure 3 which presents the cognitive components at three levels, collectively referred to as the cognitive architecture. As such, Part V begins in Section 15 with an in-depth treatment of cognitive architectures, in general, and the iCub cognitive architecture, in particular.

The iCub cognitive system is however not just a cognitive architecture. By virtue of its adherence to the developmental emergent philosophy, it requires also time and experience in order to develop its cognitive capabilities. That is, the iCub must undergo a process of ontogenesis. This process is addressed in Section 16 which deals with the experimental scenarios for the iCub's ontogenetic development.

For all the extensive mechatronic and software engineering that is required to design and build the iCub , the RobotCub project is above all else a scientific research programme in cognition. Consequently, it brings together researchers from *inter alia* the neurosciences, developmental psychology, cognitive robotics, autonomous systems theory, with the express purpose of formulating scientific hypotheses about the nature and mechanisms of cognition and testing these hypotheses on the iCub platform. The iCub cognitive architecture represents at this point an early framework in which these hypotheses can be integrated but it will no doubt change as we learn more through theoreticial and empirical research. Section 17 addresses the experimental work that we plan on conducting using the iCub . Together with the iCub cognitive architecture and the iCub's programme for ontogenesis, it represents the RobotCub consortium's research roadmap.

15 A Cognitive Architecture for the iCub

15.1 The Different Paradigms of Cognition

There are many positions on cognition, each taking a significantly different stance on the nature of cognition, what a cognitive system does, and how a cognitive system should be analyzed and synthesized. Among these, however, we can discern two broad classes: the *cognitivist* approach based on symbolic information processing representational systems, and the *emergent systems* approach, embracing connectionist systems, dynamical systems, and enactive systems, all based to a lesser or greater extent on principles of self-organization [Var92, Cla01].

Cognitivist approaches correspond to the classical and still common view that 'cognition is a type of computation' defined on symbolic representations, and that cognitive systems 'instantiate such representations physically as cognitive codes and their behaviour is a causal consequence of operations carried out on these codes' [Pyl84]. Connectionist, dynamical, and enactive systems, grouped together under the general heading of emergent systems, argue against the information processing view, a view that sees cognition as 'symbolic, rational, encapsulated, structured, and algorithmic', and argue in favour of a position that treats cognition as emergent, self-organizing, and dynamical [TS94, Kel95]. As we will see, the difference between the cognitivist and emergent positions are deep and fundamental, and go far beyond a simple distinction based on symbol manipulation. We can contrast the cognitivist and emergent paradigms on twelve distinct grounds: computational operation, representational framework, semantic grounding, temporal constraints, inter-agent epistemology, embodiment, perception, action, anticipation, adaptation, motivation, and autonomy.[5] Let us look briefly at each of these in turn.

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

Representational Framework Cognitivist systems use patterns of symbol tokens that refer to events in the external world. These are typically the descriptive[6] product of a human designer, usually, but not necessarily, punctate and local. Emergent systems representations are global system states encoded in the dynamic organization of the system's distributed network of components.

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

[5] There are many possible definitions of autonomy, ranging from the ability of a system to contribute to its own persistence [Bic00] through to the self-maintaining organizational characteristic of living creatures — dissipative far-from equilibrium systems — that enables them to use their own capacities to manage their interactions with the world, and with themselves, in order to remain viable [CH00a].

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

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

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

Embodiment Cognitivist systems do not need to be embodied, in principle, by virtue of their roots in functionalism (which states that cognition is independent of the physical platform in which it is implemented [FN99]). Emergent systems are intrinsically embodied and the physical instantiation plays a direct constitutive role in the cognitive process. [Ver06, KE06, Gar93].

Perception In cognitivist systems perception provides an interface between the external world and the symbolic representation of that world. Perception abstracts faithful spatio-temporal epresentations of the external world from sensory data. In emergent systems perception is a perturbation of the system by the environment.

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

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

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

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

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

Table 1 summarizes these points very briefly.

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

Table 1: A comparison of cognitivist and emergent paradigms of cognition; refer to the text for a full explanation.

15.2 What is a Cognitive Architecture?

Although used freely by proponents of the cognitivist, emergent, and hybrid approaches to cognitive systems, the term cognitive architecture originated with the seminal cognitivist work of Newell *et al.* [New82, New90, RLN93]. Consequently, the term has a very specific meaning in this paradigm where cognitive architectures represent attempts to create unified theories of cognition [Byr03, New90, ABB+04], *i.e.* theories that cover a broad range of cognitive issues, such as attention, memory, problem solving, decision making, learning, from several aspects including psychology, neuroscience, and computer science. Newell's Soar architecture [LNR87, RLN93, LLR98, Lew01], Anderson's ACTR architecture [And96, ABB+04], and Minsky's *Society of Mind* [Min86] are all candidate unified theories of cognition. For emergent approaches to cognition, which a focus on development from a primitive state to a fully cognitive state over the life-time of the system, the architecture of the system is equivalent to its phylogenetic configuration: the initial state from which it subsequently develops.

In the cognitivist paradigm, the focus in a cognitive architecture is on the aspects of cognition that are constant over time and that are relatively independent of the task [GYK97, RY01, Lan05]. Since cognitive architectures represent the fixed part of cognition, they cannot accomplish anything in their own right and need to be provided with or acquire knowledge to peform any given task. This combination of a given cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model*. In most cognitivist systems the knowledge incorporated into the model is normally determined by the human designer, although there is in increasing use of machine learning to augment and adapt this knowledge. The specification of a cognitive architecture consists of its representational assumptions, the characteristics of its memories, and the processes that operate on those memories. The cognitive architecture defines the manner in which a cognitive agent manages the primitive resources at its disposal [umi]. For cognitivist approaches, these resources are the computational system in which the physical symbol system is realized. The architecture specifies the formalisms for knowledge representations and the memory used to store them, the processes that act upon that knowledge, and the learning mechanisms that acquire it. Typically, it also provides a way of programming the system so that intelligent systems can be instantiated in some application domain [Lan05].

For emergent approaches, the need to identify an architecture arises from the intrinsic complexity of a cognitive system and the need to provide some form of structure within which to embed the mechanisms for perception, action, adaptation, anticipation, and motivation that enable the ontogenetic development over the system's life-time. It is this complexity that distinguishes an emergent developmental cognitive architecture from a simple connectionist robot control system that typically learns associations for specific tasks, *e.g.* the Kohonen self-organized net cited in [JV94]. In a sense, the cognitive architecture of an emergent system corresponds to the innate capabilities that are endowed by the system's phylogeny and which don't have to be learned but of course which may be developed further. There resources facilitate the system's ontogensis. They represent the initial point of departure for the cognitive system and they provide the basis and mechanism for its subsequent autonomous development, a development that may impact directly on the architecture itself. As we have stated already, the autonomy involved in this development is important because it places strong constraints on the manner in which the system's knowledge is acquired and by which its semantics are grounded (typically by autonomy-preserving anticipatory and adaptive skill construction) and by which an inter-agent epistemology is achieved (the subjective outcome of a history of shared consensual experiences among phylogenetically-compatible agents); see Table 1.

It is important to emphasize that the presence of innate capabilities in emergent systems does *not* in any way imply that the architecture is functionally modular: that the cognitive system is comprised of distinct modules each one carrying out a specialized cognitive task. If a modularity is present, it may be because it develops this modularity through experience as part of its ontogenesis or epigenesis rather than being prefigured by the phylogeny of the system (*e.g.* see Karmiloff-Smith's theory of representational redescription, [KS92, KS94]). Even more important, it does not necessarily imply that the innate capabilities are hard-wired cognitive skills as suggested by nativist psychology (*e.g.* see Fodor [Fod83] and Pinker [Pin97]).[7] At the same time, neither does it necessarily imply that the cognitive system is a blank slate, devoid of any innate cognitive structures as posited in Piaget's constructivist view of cognitive development [Pia55];[8] at the very least there must exist a mechanism, structure, and organization which allows the cognitive system to be autonomous, to act effectively to some limited extent, and to develop that autonomy.

Finally, since the emergent paradigm sits in opposition to the two pillars of cognitivism—the dualism that posits the logical separation of mind and body, and the functionalism that posits that cognitive mechanisms are independent of the physical platform [FN99] — it is likely that the architecture will reflect or recognize in some way the morphology of the physical body of which it is embedded and of which it is an intrinsic part.

Having established these boundary conditions for cognitivist and emergent cognitive architectures (and implicitly for hybrid architectures), for the purposes of this review the term cognitive architecture

[7] More recently, Fodor [Fod00] asserts that modularity applies only to local cognition (*e.g.* recognizing a picture of Mount Whitney) but not global cognition (*e.g.* deciding to trek the John Muir Trail).

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

will the taken in the general and non-specific sense. By this we mean the minimal configuration of a system that is necessary for the system to exhibit cognitive capabilities and behaviours: the specification of the components in a cognitive system, their function, and their organization as a whole. That said, since the RobotCub project is committed to the emergent paradigm, we do place particular emphasis on the need of systems that are developmental and emergent, rather than pre-configured.

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

Table 2: The cognitive architectures reviewed in this section.

Below, we will review several cognitive architectures drawn from the cognitivist, emergent, and hybrid traditions, beginning with some of the best known cognitivist ones. Table 2 lists the cognitive architectures reviewed under each of these three headings. Following this review, we present a comparative analysis of these architectures using a subset of the twelve paradigm characteristics we discussed in Section 15.1: computational operation, representational framework, semantic grounding, temporal constraints, inter-agent epistemology, role of physical instantiation, perception, action, anticipation, adaptation, motivation, embodiment, autonomy.

15.3 A Review of Cognitive Architectures

15.3.1 The Soar Cognitive Architecture

The Soar system [LNR87, RLN93, LLR98, Lew01] is Newell’s candidate for a Unified Theory of Cognition [New90]. It is a production (or rule-based) system[9] that operates in a cyclic manner, with a production cycle and a decision cycle. It operates as follows. First, all productions that match the contents of declarative (working) memory fire. A production that fires may alter the state of declarative memory and cause other productions to fire. This continues until no more productions fire. At this point, the decision cycle begins in which a single action from several possible actions is selected. The selection is based on stored action preferences. Thus, for each decision cycle there may have been many production cycles. Productions in Soar are low-level; that is to say, knowledge is encapsulated at a very small grain size.

One important aspect of the decision process concerns a process known as *universal sub-goaling*. Since there is no guarantee that the action preferences will be unambiguous or that they will lead to a unique action or indeed any action, the decision cycle may lead to an ‘impasse’. If this happens, Soar sets up an new state in a new problem space — sub-goaling — with the goal of resolving the impasse. Resolving one impasse may cause others and the sub-goaling process continues. It is assumed that degenerate cases can be dealt with (*e.g.* if all else fails, choose randomly between two actions). Whenever an impasse is resolved, Soar creates a new production rule which summarizes the

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

processing that occurred in the sub-state in solving the sub-goal. Thus, resolving an impasse alters the system super-state, *i.e.* the state in which the impasse originally occurred. This change is called a result and becomes the outcome of the production rule. The condition for the production rule to fire is derived from a dependency analysis: finding what declarative memory items matched in the course of determining the result. This change in state is a form of learning and it is the only form that occurs in Soar, *i.e.* Soar only learns new production rules. Since impasses occur often in Soar, learning is pervasive in Soar's operation.

15.3.2 EPIC— Executive Process Interactive Control

EPIC [KM97] is a cognitive architecture that was designed to link high-fidelity models of perception and motor mechanisms with a production system. An EPIC model requires both knowledge encapsulated in production rules and perceptual-motor parameters. There are two types of parameter: standard or system parameters which are fixed for all tasks (*e.g.* the duration of a production cycle in the cognitive processor: 50 ms) and typical parameters which have conventional values but can vary between tasks (*e.g.* the time required to effect recognition of shape by the visual processor: 250 ms).

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

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

Like Soar, the cognitive processor in EPIC is a production system in which multiple rules can fire in one production cycle. However, the productions in EPIC have a much larger grain size than Soar productions.

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

EPIC does not have any learning mechanism.

15.3.3 ACT-R— Adaptive Control of Thought – Rational

The ACT-R [And96, ABB+04] cognitive architecture is another approach to creating an unified theory of cognition. It focusses on the modular decomposition of cognition and offers a theory of how these modules are integrated to produce coherent cognition. The architecture comprises five specialized modules, each devoted to processing a different kind of information (see Figure 4). 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.

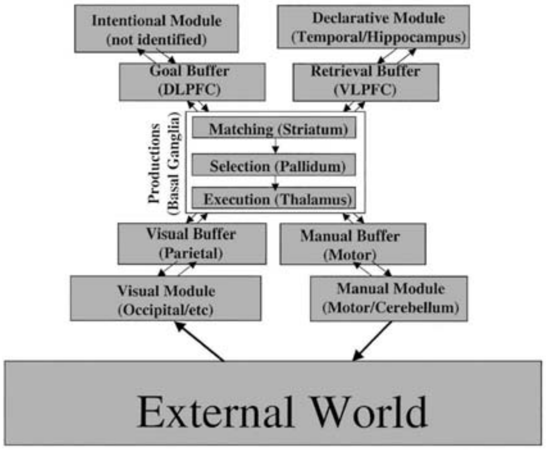

Intentional Module (not identified)
Declarative Module (Temporal/Hippocampus)
Goal Buffer (DLPFC)
Retrieval Buffer (VLPFC)
Productions (Basal Ganglia)
Matching (Striatum)
Selection (Pallidum)
Execution (Thalamus)
Visual Buffer (Parietal)
Manual Buffer (Motor)
Visual Module (Occipital/etc)
Manual Module (Motor/Cerebellum)
External World

Figure 4: The ACT-R Cognitive Architecture (from [ABB+04]).

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

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

Whilst early incarnations of ACT-R focussed primarily on the production system, the importance of perceptuo-motor processes in determining the nature of cognition is recognized by Anderson *et al.* in more recent versions [Byr03, ABB+04]. That said, the perceptuo-motor system in ACT-R is based on the EPIC architecture [KM97] 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 [And96]. 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.

The ACT-R visual module differs somewhat from the EPIC visual system in that it is separated into two sub-modules, each with its own buffer, one for object localization and associated with the dorsal pathway, and the other for object recognition and associated with the ventral pathway. Note that this sharp separation of function between the ventral and dorsal pathways has been challenged by recent neurophysiological evidence which points to the interdependence between the two pathways [RFG97, RFFG97]. When the production system requests information from the localization module, it can supply constraints in the form of attribute-value pairs (*e.g.* colour-red) and the localization module will then place a chunk in its buffer with the location of some object that satisfies those constraints.

The production system queries the recognition system by placing a chunk with location information in its buffer; this causes the visual system to subsequently place a chunk representing the object at that location in its buffer for subsequent processing by the production system. This is a significant idealization of the perceptual process.

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

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

Procedural memory is encapsulated in the production system which coordinates the overall operation of the architecture. Whilst several productions may qualify to fire, only one production is selected. This selection is called conflict resolution. The production selected is the one with the highest utility, a factor which is a function of an estimate of the probability that the current goal will be achieved if this production is selected, the value of the current goal, and an estimate of the cost of selecting the production (typically proportional to time), both of which are learned in a Bayesian framework from previous experience with that production. In this way, ACT-R can adapt to changing circumstances [Byr03].

Declarative knowledge effectively encodes things in the environment while procedural knowledge encodes observed transformations; complex cognition arises from the interaction of declarative and procedureal knowledge [And96]. A central feature of the ACT-R cognitive architecture is that these two types of knowledge are tuned in specific application by encoding the statistics of knowledge. Thus, ACT-R learns sub-symbolic information by adjusting or tuning the knowledge parameters. This sub-symbolic learning distiguishes ACT-R from the symbolic (production-rule) learning of Soar.

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

Like Soar, ACT-R has evolved significantly over several years [And96]. It is currently in Version 5.0 [ABB+04].

15.3.4 The ICARUS Cognitive Architecture

The ICARUS cognitive architecture [Lan04, Lan05, CKL+04, Lan06] follows in the tradition of other cognitivist architectures, such ACT-R, Soar, and EPIC, exploiting symbolic representations of knowledge, the use of pattern matching to select relevant knowledge elements, operation according to the conventional recognize-act cycle, and an incremental approach to learning. In this, ICARUS is adheres strictly to the Newell and Simon's physical symbol system hypothesis [NS76] which states that symbolic processing is a necessary and sufficient condition for intelligent behaviour. However, ICARUS goes further and claims that mental states are always grounded in either real or imagined physical states, and *vice versa* that problem-space symbolic operators always expand to actions that can be effected or executed. Langley refers to this as the *symbolic physical system* hypothesis. This assertion of the importance of action and perception is similar to recent claims by others in the cognitivist community such as Anderson *et al.* [ABB+04].

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

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

15.3.5 ADAPT—A Cognitive Architecture for Robotics

Some authors, e.g. Benjamin *et al.* [BLL04], argue that existing cognitivist cognitive architectures such as Soar, ACT-R, and EPIC, don't easily support certain mainstream robotics paradigms such as adaptive dynamics and active perception. Many robot programs comprise several concurrent distributed communicating real-time behaviours and consequently these architectures are not suited since their focus is primarily on "sequential search and selection", their learning mechanisms focus on composing sequential rather than concurrent actions, and they tend to be hierarchically-organized rather than distributed. Benjamin *et al.* don't suggest that you cannot address such issues with these architectures but that they are not central features. They present a different cognitive architecture, ADAPT— Adaptive Dynamics and Active Perception for Thought, which is based on Soar but also adopts features from ACT-R (such as long-term declarative memory in which sensori-motor schemas to control perception and action are stored) and EPIC (all the perceptual processes fire in parallel) but the low-level sensory data is placed in short-term working memory where it is processed by the cognitive mechanism. ADAPT has two types of goals: task goals (such as 'find the blue object') and architecture goals (such as 'start a schema to scan the scene'). It also has two types of actions: task actions (such as 'pick up the blue object') and architectural actions (such as 'initiate a grasp schema'). While the architectural part is restricted to allow only one goal or action at any one time, the task part has no such restrictions and many task goals and actions — schemas — can be operational at the same time. The architectural goals and actions are represented procedurally (with productions) while the task goals and actions are represented declaratively in working memory as well as procedurally.

15.3.6 Autonomous Agent Robotics

Autonomous agent robotics (AAR) and behaviour-based systems represents an emergent alternative to cognitivist approaches. Instead of a cognitive system architecture that is based on a decomposition into functional components (*e.g.* representation, concept formation, reasoning), an AAR architecture is based on interacting *whole* systems. Beginning with simple whole systems that can act effectively in simple circumstances, layers of more sophisticated systems are added incrementally, each layer subsuming the layers beneath it. This is the subsumption architecture introduced by Brooks [Bro86]. Christensen and Hooker [CH00b] argue that AAR is not sufficient either as a principled foundation for a general theory of situated cognition. One limitation includes the explosion of systems states that results from the incremental integration of sub-systems and the consequent difficulty in coming up with an initial well-tuned design to produce coordinated activity. This in turn imposed a need from some form of self-management, something not included in the scope of the original subsumption architecture. A second limitation is that it becomes increasingly problematic to rely on environmental cues to achieve the right sequence of actions or activities as the complexity of the task rises. AAR is also insufficient for the creation of a comprehensive theory of cognition: as the subsumption architecture can't be scaled to provide higher-order cognitive faculties (it can't explain self-directed behaviour) and even though the behaviour of an AAR system may be very complex it is still ultimately a reactive system.

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

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

15.3.7 A Global Workspace Cognitive Architecture

Shanahan [Sha06, SB05, Sha05b, Sha05a] proposes a biologically-plausible brain-inspired neurallevel cognitive architecture in which cognitive functions such as anticipation and planning are realized through internal simulation of interaction with the environment. Action selection, both actual and internally-simulated, is mediated by affect. The architecture is based on an external sensori-motor loop and an internal sensori-motor loop in which information passes though multiple competing cortical areas and a global workspace.

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

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

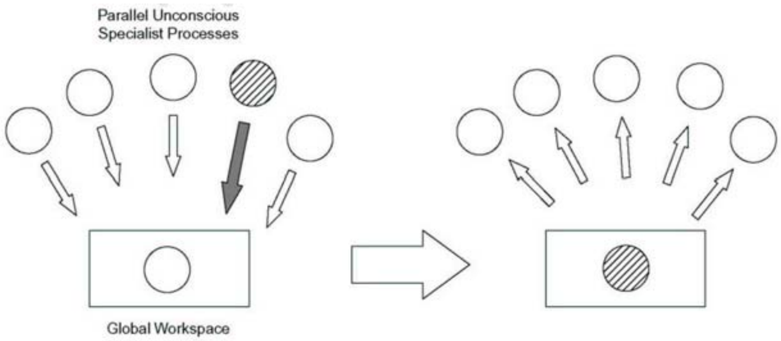


Figure 5: The Global Workspace Theory cognitive architecture: ’winner-take-all’ coordination of competing concurrent processes (from [Sha06]).

Shanahan’s cognitive architecture is founded also upon the fundamental importance of parallelism as a constituent component in the cognitive process as opposed to being a mere implementation issue. He deploys the *global workspace* model [Baa98, Baa02] of information flow in which a sequence of states emerges from the interaction of many separate parallel processes (see Figure 5). These specialist processes compete and co-operate for access to a global workspace. The winner(s) of the competition gain(s) controlling access to the global access and can then broadcast information back to the competing specialist processes. Shanahan argues that this type of architecture provides an elegant solution to the frame problem.

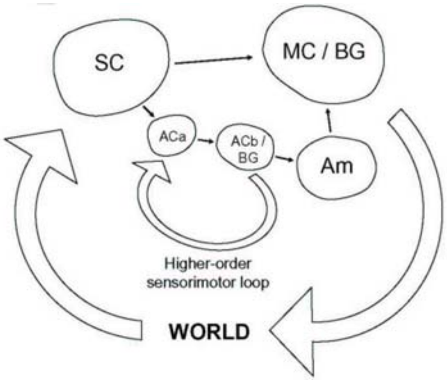


Figure 6: The Global Workspace Theory cognitive architecture: achieving prospection by sensori-motor simulation (from [Sha06]).

Shanahan’s cognitive architecture is comprised of the following components: a first-order sensorimotor loop, closed externally through the world, and a higher-order sensori-motor loop, closed internally through associative memories (see Figure 5). The first-order loop comprises the sensory cortex and the basal ganglia (controlling the motor cortex), together providing a reactive action-selection sub-system. The second-order loop comprises two associative cortex elements which carry out offline simulations of the system’s sensory and motor behaviour, respectively. The first associative cortex simulates a motor output while the second simulates the sensory stimulus expected to follow from a given motor output. The higher-order loop effectively modulates basal ganglia action selection in the first-order loop via an affect-driven amygdala component. Thus, this cognitive architecture is able to anticipate and plan for potential behaviour through the exercise of its “imagination” (*i.e.* its associative

internal sensori-motor simulation. The global workspace doesn't correspond to any particular localized cortical area. Rather, it is a global communications network.
The architecture is implemented as a connectionist system using G-RAMs: generalized random access memories [Ale90]. Interpreting its operation in a dynamical framework, the global workspace and competing cortical assemblies each define an attractor landscape. The perceptual categories constitute attactors in a state space that reflects the structure of the raw sensory data. Prediction is achieved by allowing the higher-order sensori-motor loop to traverse along a simulated trajectory in that state space so that the global workspace visits a sequence of attractors. The system has been validated in a Webot [Mic04] simulation environment.

15.3.8 Self-Directed Anticipative Learning

Christensen and Hooker propose a new emergent interactivist-constructivist (I-C) approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [CH00a]. This approach falls under the broad heading of dynamical embodied approaches in the non-cognitivist paradigm. They assert first the primary model for cognitive learning is anticipative skill construction and that processes that both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems. For them, intelligence is a continuous management process that has to support the need to achieve autonomy in a living agent, distributed dynamical organization, and the need to produce functionally coherent activity complexes that match the constraints of autonomy with the appropriate organization of the environment across space and time through interaction. In presenting their approach they use the term "explicit norm signals" for the signals that a system uses to differentiate an appropriate context performing an action. These norm signals reflect conditions for the (maintenance) of the system's autonomy (*e.g.* hunger signals depleted nutritional levels). The complete set of norm signals is termed the norm matrix. They then distinguish between two levels of management: low-order and high-order. Low-order management employs norm signals which differentiate only a narrow band of the overall interaction process of the system (*e.g.* a mosquito uses heat tracking and gradient tracking to seek blood hosts). Since it uses only a small number of parameters to direct action, success ultimately depends on simple regularity in the environment. These parameters also tend to be localized in time and space. On the other hand, high-order management strategies still depend to an extent on regularity in the environment but exploit parameters that are more extended in time and space and use more aspects of the interactive process, including the capacity to anticipate and evaluate the system's performance, to produce effective action (and improve performance). This is the essence of self-directedness. "Self-directed systems anticipate and evaluate the interaction process and modulate system action accordingly". The major features of selfdirectedness are action modulation ("generating the right kind of extended interaction sequences"), anticipation ("who will/should the interaction go?", evaluation ("how did the evaluation go?"), and constructive gradient tracking ("learning to improve performance").

15.3.9 A Self-Affecting Self-Aware (SASE) Cognitive Architecture

Weng [Wen04a, Wen04b, Wen02] introduced an emergent cognitive architecture that is specifically focussed on the issue of development by which he means that the processing accomplished by the architecture is not specified (or programmed) *a priori* but is the result of the real-time interaction of the system with the environment including humans. Thus, the architecture is not specific to tasks, which are unknown when the architecture is created or programmed, but is capable of adapting and developing to learn both the tasks required of it and the manner in which to achieve the tasks. In this sense, even through Weng's architecture is not a cognitivist one, his use of the term is very faithful to the meaning of *cognitive architecture* as it was originally intended when it was introduced originally in the cognitivist paradigm. That is, it represents the underlying infrastructure for a cognitive system, specifically those aspects of a cognitive agent that are constant over time and independent of the task [RY01, GYK97, Lan05].

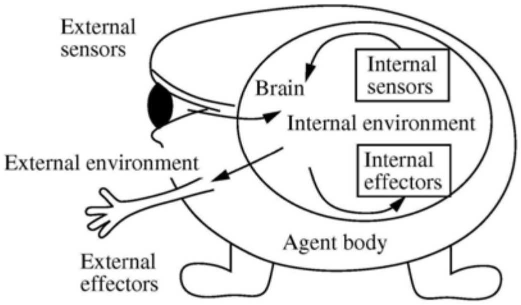


Figure 7: The Self-Aware Self-Effecting (SASE) architecture (from [Wen02]).

Weng refers to his architecture as a Self-Aware Self-Effecting (SASE) system (see Figure 7). The architecture entails an important distinction between the sensors and effectors that are associated with the environment (including the system's body and thereby including proprioceptive sensing) and those that are associated with the system's 'brain' or central nervous system (CNS). Only those systems that have explicit mechanisms for sensing and affecting the CNS qualify as SASE architectures. The implications for development are significant: the SASE architecture is configured with no knowledge of the tasks it will ultimately have to perform, its brain or CNS are not directly accessible to the (human) designers once it is launched, and after that the only way a human can affect the agent is through the external sensors and effectors. Thus, the SASE architecture is very faithful to the emergent paradigms of cognition, especially the enactive approach: its phylogeny is fixed and it is only through ontogenetic development that the system can learn to operate effectively in its environment.

The concept of self-aware self-effecting operation is similar to the level 2 autopoietic organizational principles introduced by Matura and Varela [MV87] (*i.e.* both self-production and self-development) and is reminiscent of the recursive self-maintenant systems principles of Bickhard [Bic00] and Christensen's and Hooker's interactivist-constructivist approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [CH00a]. Weng's contribution differs in that he provides a specific computational framework in which to implement the architecture. Weng's cognitive architecture is based on Markov Decision Processes (MDP), specifically a developmental observation-driven self-aware self-effecting Markov Decision Process (DOSASE MDP). Weng places this particular architecture in a spectrum of MDPs of varying degrees of behavioural and cognitive complexity [Wen04b]; the DOSASE MDP is type 5 of six different types of architecture and is the first type in the spectrum that provides for a developmental capacity. Type 6 builds on this to provide additional attributes, specifically greater abstraction, self-generated contexts, and a higher degree of sensory integration.

The example DOSASE MDP vision system detailed in [Wen04a] further elaborates on the cognitive architecture, detailing three types of mapping in the information flow within the architecture: sensory mapping, cognitive mapping, and motor mapping. It is significant that there is more than one cognitive pathway between the sensory mapping and the motor mapping, one of which encapsulates innate behaviours (and the phylogenetically-endowed capabilities of the system) while the other encapsulates learned behaviours (and the ontogenetically-developed capabilities of the system). These two pathways are mediated by a subsumption-based motor mapping which accords higher priority to the ontogenetically-developed pathway. A second significant feature of the architecture is that it facilitates what Weng refers to as "primed sensations" and "primed action". These correspond to predictive sensations and actions and thereby provide the system with the anticipative and prospective capabilities that are the hallmark of cognition.

The general SASE schema, including the associated concept of Autonomous Mental Development (AMD), has been developed and validated in the context of two autonomous developmental robotics systems, SAIL and DAV [WHZ+00, WZ02, Wen04a, Wen04b].

15.3.10 Darwin: Neuromimetic Robotic Brain-Based Devices

Kirchmar *et al.* [KE05, KNGE05, KSN+05, KR05, KE06, SME+04] have developed a series of robot platforms called Darwin to experiment with developmental agents. These systems are 'brain-based devices' BBDs which that exploit a simulated nervous system that can develop spatial and episodic memory as well as recognition capabilities through autonomous experiential learning. As such, BDDs are a neuromimetic approach in the emergent paradigm that is most closely aligned with the enactive and the connectionist models. It differs from most connectist approaches in that the architecture is much more strongly modelled on the structure and organization of the brain than are conventional artificial neural networks, *i.e.* they focus on the nervous system as a whole, its constituent parts, and their interaction, rather than on a neural implementation of some individual memory, control, or recognition function.

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

Darwin VIII is capable of discriminating reasonably simple visual targets (coloured geometric shapes) by associating it with an innately preferred auditory cue. Its simulated nervous system contains 28 neural areas, approximately 54,000 neuronal units, and approximately 1.7 million synaptic connections. The architecture comprises regions for vision (V1, V2, V4, IT), tracking (C), value or saliency (S), and audition (A). Gabor filtered images, with vertical, horizontal, and diagonal selectivity, and red-green colour filters with on-centre off-surround and off-centre on-surround receptive fields, are fed to V1. Sub-regions of V1 project topographically to V2 which in turn projects to V4. Both V2 and V4 have excitatory and inhibitory reentrant connections. V4 also has a non-topographical projection back to V2 as well as a non-topographical projection to IT, which itself has reentrant adaptive connections. IT also projects non-toographically back to V4. The tracking area (C) determines the gaze direction of Darwin VIII's camera based on excitatory projections from the auditory region A. This causes Darwin to orient toward a sound source. V4 also projects topographically to C causing Darwin VIII to centre its gaze on a visual object. Both IT and the value system S have adaptive connections to C which facilitates the learned target selection. Adaptation is effected using the Hebbian-like Bienenstock-Cooper-Munroe (BCM) rule [BCM82]. From a behavioural perspective, Darwin VIII is conditioned to prefer one target over others by associating it with the innately peferred auditory cue and to demonstrate this preference by orienting towards the target.

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

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

15.3.11 A Humanoid Robot Cognitive Architecture

Burghart *et al.* [BMS+05] present a hybrid cognitive architecture for a humanoid robot. It is based on interacting parallel behaviour-based components, comprising a three-level hierarchical perception sub-system, a three-level hierarchical task handling system, a long-term memory sub-system based on a global knowledge database (utilizing a variety of representational schemas, including object ontologies and geometric models, Hidden Markov Models, and kinematic models), a dialogue manager which mediates between perception and task planning, an execution supervisor, and an 'active models' short-term memory sub-system to which all levels of perception and task management have access. These active models play a central role in the cognitive architecture: they are initialized by the global knowledge database and updated by the perceptual sub-system and can be autonomously actualized and reorganized. The perception sub-system comprises a three-level hierarchy with low, mid, and high level perception modules. The low-level perception module provides sensor data interpretation without accessing the central system knowledge database, typically to provide reflex-like low-level robot control. It communicates with both the mid-level perception module and the task execution module. The mid-level perception module provides a variety of recognition components and communicates with both the system knowledge database (long-term memory) as well as the active models (short-term memory). The high-level perception module provides more sophisticated interpretation facilities such as situation recognition, gesture interpretation, movement interpretation, and intention prediction.

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

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

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

15.3.12 The Cerebus Architecture

Horswill [Hor01, Hor06] argues that classical artificial intelligence systems such as those in the tradition of Soar, ART-R, and EPIC, are not well suited for use with robots. Traditional systems typically store all knowledge centrally in a symbolic database of logical assertions and reasoning is concerned mainly with searching and sequentially updating that database. However, robots are distributed systems with multiple sensory, reasoning, and motor control proceses all running in parallel and often only loosely coupled with one another. Each of these processes maintains its own separate and limited representation of the world and the task at hand and he argues that it is not realistic to require them to constantly synchronize with a central knowledge base.

Recently, much the same argument has been made by neuroscientists about the structure and operation of the brain. For example, evidence suggest that space perception is not the result of a single circuit, and in fact derives from the joint activity of several fronto-parietal circuits, each of which encodes the spatial location and transforms it into a potential action in a distinct and motor-specific manner [RFG97, RFFG97]. In other words, the brain encodes space not in a single unified manner —there is no general purpose space map —but in many different ways, each of which is specifically concerned with a particular motor goal. Different motor effectors need different sensory input: derived in different ways and differently encoded in ways that are particular to the different effectors. Conscious space perception emerges from these different pre-existing spatial maps.

Horswill contends also that the classical reasoning systems don't have any good way of directing perceptual attention: they either assume that all the relevant information is already stored in the database or they provide a set of actions that fire task-specific perceptual operators to update specific parts of the database (just as, for example, happens in ACT-R). Both of these approaches are problematic: the former fall foul of the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new data) and the second requires that the programmer design the rule based to ensure that the appropriate actions are fired in the right circumstances and at the right time; see also similar arguments by Christensen and Hooker [CH00b].

Horswill argues that keeping all of the distinct models or representations in the distributed processes or sub-systems consistent needs to be a key focus of the overall architecture and that is should be done without sychronizing with a central knowledge base. They propose a hybrid cognitive architecture, Cerebus, that combines the tenets of behaviour-based architectures with some features of symbolic AI (forward- and backward-chaining inference using predicate logic). It represents an attempt to scale behaviour-based robots (*e.g.* see Brooks [Bro86] and Arkin [Ark98]) 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 longterm 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.

15.3.13 Cog: Theory of Mind

Cog [BBM+99] is an upper-torso humanoid robot platform for research on developmental robotics. Cog has a pair of six degree-of-freedom arms, a three degree-of-freedom torso, and a seven degreeof-freedom head and neck. It has a narrow and wide angle binocular vision system (comprising four colour cameras), an auditory system with two microphones, a three-degree of freedom vestibular system, and a range of haptic sensors.

As part of this project, Scassellati has put forward a proposal for a Theory of Mind for Cog [Sca02] that focusses on social interaction as a key aspect of cognitive function in that social skills require the attribution of beliefs, goals, and desires to other people.

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

Scassellati's proposed architecture is based on Leslie's model of Theory of Mind [Les94] and Baron-Cohen's model of Theory of Mind [BC95] both of which decompose the problem into sets of precursor skills and developmental modules, albeit in a different manner. Leslie's Theory of Mind emphasizes independent domain specific modules to distinguish (a) mechanical agency, (b) actional agency, and (c) attitudinal agency; roughly speaking the behaviour of inanimate objects, the behaviour of animate objects, and the beliefs and intentions of animate objects. Baron-Cohen's Theory of Mind comprises three four modules, one of which is concerned with the interpretation of perceptual stimuli (visual, auditory, and tactile) associated with self-propelled motion, and one of which is concerned with the interpretation of visual stimuli associated with eye-like shapes. Both of these feed a shared attention module which in turn feed a Theory of Mind module that represents intentional knowledge or 'epistemic mental states' of other agents.

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

15.3.14 Kismet

The role of emotion and expressive behaviour in regulating social interaction between humans and robots has been examined by Breazeal using an expressive articulated anthropomorphic robotic head called Kismet [Bre00, Bre03]. Kismet has a total of 21 degree-of-freedom, three to control the head orientation, three to direct the gaze, and fifteen to control the robots facial features (*e.g.* eye-lids, eyebrows, lips, and ears). Kismet has a narrow and wide angle binocular vision system (comprising four colour cameras), and two microphones, one mounted in each ear. Kismet is designed to engage people in natural and expressive face-to-face interaction, perceiving a natural social cues and responding through gaze direction, facial expression, body posture, and vocal babbling.

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

Kismet has two types of motivations: drives and emotions. Drives establish the top-level goals of the robot: to engage people (social drive), to engage toys (stimulation drive), and to occasionally rest (fatigue drive). The robot's behaviour is focussed on satiating its drives. These drives have a longer time constant compared with emotions. and they operate cyclically: increasing in the absence of satisfying interaction and diminishing with habituation. The goal is to keep the drive level somewhere in a homeostatic region between under stimulation and over stimulation. Emotions — anger & frustration, disgust, fear & distress, calm, joy, sorrow, surprise, interest, boredom — elicit specific behavioural responses such as complain, withdraw, escape, display pleasure, display sorrow, display startled response, re-orient, and seek, in effect tending to cause the robot to come into contact with things that promote its "well-being" and avoid those that don't. Emotions are triggered by pre-specified antecedent conditions which are based on perceptual stimuli as well as the current drive state and behavioural state.

Kismet has five distinct modules in its cognitive architecture: a perceptual system, an emotion system, a behaviour system, a drive system, and a motor system (see Figure 8).

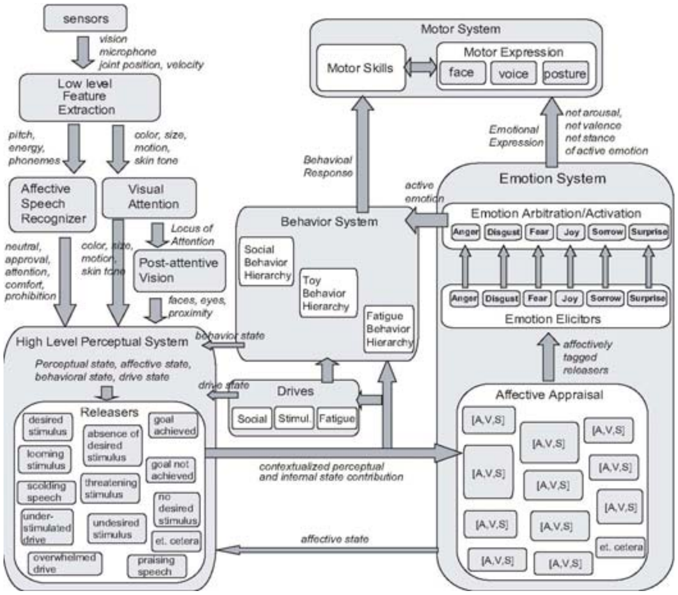


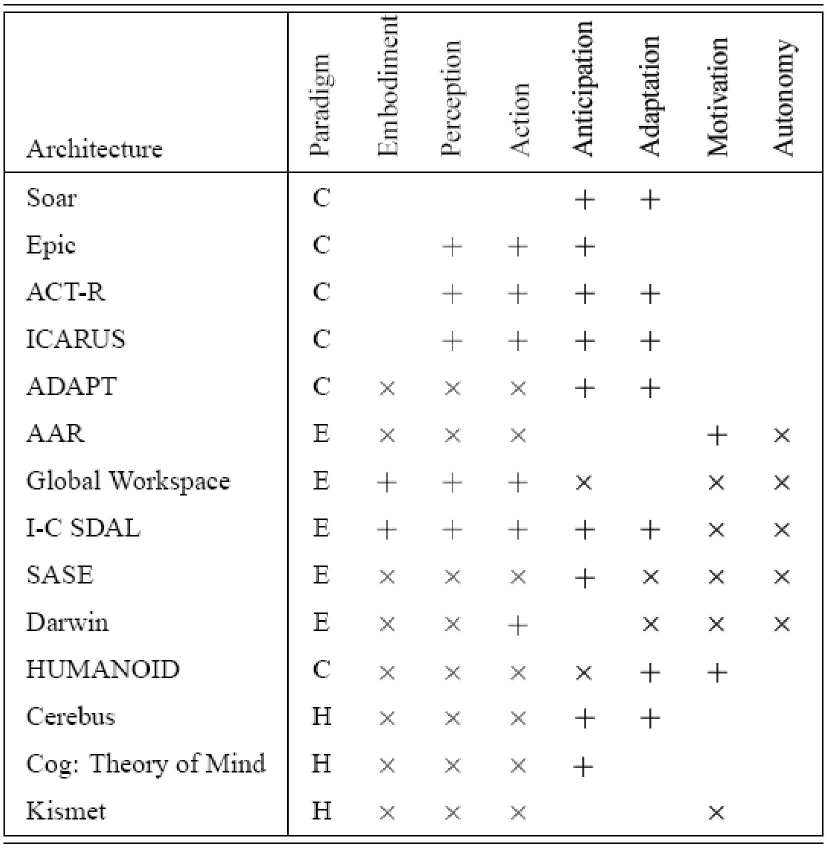
Figure 8: The Kismet cognitive architecture (from [Bre03]).

The perceptual system comprises a set of low-level processes which sense visual and auditory stimuli, perform feature extraction (*e.g.* colour, motion, frequency), extract affective descriptions from speech, orient visual attention, and localize relevant features such as faces, eyes, objects, *etc.*. These are input to a high level perceptual system where, together with affective input from the emotion system, input from the drive system and the behaviour system, they are bound by *releaser* processes 'that encode the robot's current set of beliefs about the state of the robot and its relation to the world. There are many different kinds of releasers, each of which is 'hand-crafted' by the system designer. When the activation level of a releaser exceeds a given threshold (based on the perceptual, affective, drive, and behavioural inputs) it is output to the emotion system for appraisal. Breazeal says that 'each releaser can be thought of as a simple "cognitive" assessment that combines lower-level perceptual features with measures of its internal state into behaviorally significant perceptual categories' [Bre03]. The appraisal process tags the releaser output with pre-specified (*i.e.* designed-in) affective information on their arousal (how much it stimulates the system), valence (how much it is favoured), and stance (how approachable it is). These are then filtered by 'emotion elicitor' to map each AVS (arousal, valence, stance) triple onto the individual emotions. A single emotion is then selected by a winner-take-all arbitration process, and output to the behaviour system and the motor system to evoke the appropriate expression and posture.

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

15.4 Comparison

Table 3 shows a summary of all the architectures reviewed *vis-à-vis* a subset of the twelve characteristics of cognitive systems which we discussed in Chapter 15.1. We have omitted the first five characteristics — Computation Operation, Representational Framework, Semantic Grounding, Temporal Constraints, and Inter-agent Epistemology — because these can be inferred directly by the paradigm in which the system is based: cognitivist, emergent, or hybrid, denoted by a C, E, or H in Table 3. A '×' indicates that the characteristic is strongly addressed in the architecture, '+' indicates that it is weakly addressed, and a space indicates that it is not addressed at all in any substantial manner. A '×' is assigned under the heading of Adaptation only if the system is capable of development (in the sense of creating new representational frameworks or models) rather than simple learning (in the sense of model parameter estimation) [Wen04a].



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

Table 3: Cognitive architections *vis-à -vis* the seven of the twelve characteristics of cognitive systems. Key: '×' indicates that the characteristic is strongly addressed in the architecture, '+' indicates that it is weakly addressed, and a space indicates that it is not addressed at all in any substantial manner. A '×' is assigned under the heading of Adaptation only if the system is capable of development (in the sense of creating new representational frameworks or models) rather than simple learning (in the sense of model parameter estimation). C, E, and H denote cognitivist, emergent, and hybrid paradigms, respectively.

15.5 Implications for the Development of Cognition in Artificial Systems

We finish this survey by drawing together the main issues raised in the foregoing and we summarize some of the key features that a system capable of autonomous mental development, *i.e.* an artificial cognitive system, should exhibit, especially those that adhere to a developmental approach.

Krichmar *et al.* identify six design principles for systems that are capable of development [KE05, KR05, KE06]. Although they present these principles in the context of their brain-based devices, most are directly applicable to emergent systems in general. First, they suggest that the architecture should address the dynamics of the neural element in different regions of the brain, the structure of these regions, and especially the connectivity and interaction between these regions. Second, they note that the system should be able to effect perceptual categorization: *i.e.* to organize unlabelled sensory signals of all modalities into categories without *a priori* knowledge or external instruction. In effect, this means that the system should be autonomous and, as noted by Weng [Wen04a], p. 206, a developmental system should be a model generator, rather than a model fitter (*e.g.* see [ONP06]). Third, a developmental system should have a physical instatiation, *i.e.* it should be embodied, so that it is tightly coupled with its own morphology and so that it can explore its environment. Fourth, the system should engage in some behavioural task and, consequently, it should have some minimal set of innate behaviours or reflexes in order to explore and survive in its initial environmental niche.

From this minimum set, the system can learn and adapt so that it improves[10] its behaviour over time. Fifth, developmental systems should have a means to adapt. This implies the presence of a value system (*i.e.* a set of motivations that guide or govern it development). These should be non-specific (in the sense that they don't specify what actions to take) modulatory signals that bias the dynamics of the system so that the global needs of the system are satisfied: in effect, so that its autonomy is preserved or enhanced. Such value systems might possibly be modelled on the value system of the brain: dopaminergic, cholinergic, and noradrenergic systems signalling, on the basis of sensory stimuli, reward prediction, uncertainty, and novelty. Krichmar *et al.* also note that brain-based devices should lend themselves to comparison with biological systems.

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

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

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

Third, and because cognitive systems are not only adaptive but also anticipatory and prospective, it is crucial that they have (by virtue of their phylogeny) or develop (by virtue of their ontogeny)

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

some mechanism to rehearse hypothetical scenarios —explicitly like Anderson's ACT-R architecture [ABB+04] or implicitly like Shanahan's global workspace dynamical architecture [Sha06] — and a mechanism to then use this to modulate the actual behaviour of the system.

Finally, developmental cognitive systems have to be embodied, at the very least in the sense of stuctural coupling with the environment and probably in some stronger organismoid form [Zie01, Zie03], if the epistemological understanding of the developed systems is required to be consistent with that of other cognitive agents such as humans [Ver06]. 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. It is for this reason that the iCub has been equipped with 53 degrees of freedom to effect looking, locomotion by crawling, sitting, reaching, grasping, dexterous manipulation, imitation, and social interaction.

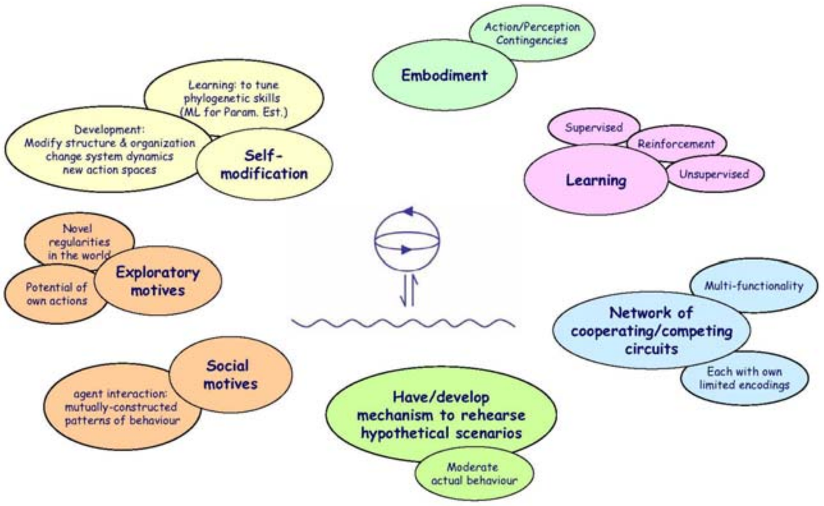

Action/Perception Contingencies
Embodiment
Learning: to tune phylogenetic skills (ML for Param. Est.)
Development: Modify structure & organization change system dynamics new action spaces
Self-modification
Supervised
Reinforcement
Learning
Unsupervised
Novel regularities in the world
Exploratory motives
Potential of own actions
Multi-functionality
Network of cooperating/competing circuits
Each with own limited encodings
Social motives
agent interaction: mutually-constructed patterns of behaviour
Have/develop mechanism to rehearse hypothetical scenarios
Moderate actual behaviour

Figure 9: Requirements for a developmental cognitive architecture.

These requirements are summarized in Figure 9. We reiterate them again here in list form for ease of reference.

Structure

1. Confederation of competing and cooperating distributed multi-functional sub-systems (cortical circuits);
2. Each with its own limited encoding/representational framework;
3. Each effecting a given phylogenetic ability;
4. Each exploiting some form of self-organization;
5. Together achieving the goal of effective behaviour: this implies some form of modulation, effected either by some self-synchronizing mechanism or by some modulation circuit.

Anticipation & Prospection

1. Mechanism to rehearse hypothetical scenarios;
2. Mechanism to facilitate modulation of perceptuo-motor behaviours.

Adaptivity & Self-Modification

1. Parameter adjustment of phylogentic skills through learning;
2. Self-modification of the system structure and/or organization:

- to alter the system dynamics based on experience;
- to expand the iCub's repertoire of actions;
- while effecting homeostasis of overall system organization;

3. Motivations for development:

- exploratory drives to discovery of novel regularities in perceptuo-motor space;
- social drives to create mutually-constructed patterns of behaviour through shared activities.

4. Modes of Learning:

- Supervised learning;
- Reinforcement learning;
- Unsupervised learning.

15.6 The iCub Cognitive Architecture

Having identified the general requirements and principles of a developmentally-based, emergent, embodied, artificial cognitive system, we now proceed to say how exactly these requirements can be satisfied and exploit these principles in creating the iCub cognitive architecture.

In what follows, we present a proposal for a cognitive architecture. In its current state, this is very much a strawman architecture: it needs to be validated and, inevitably, it will need to be revised and amended as the project progresses. This validation should be both empirical (through experiment) and theoretical (through reference to neuroscientific and psychological models). Its role at the present is primarily to exercise the partners working hypotheses on cognition and to act as a mechanism to drive the development of functioning cognitive software for the iCub, implementing both phylogentic and ontogenetic processes. It is worth remarking that the cognitive architecture also needs to be compatible with the software architecture which will be used to facilitate this implementation, and, consequently, it needs to facilitate open and easy usage, in part or as a whole, by any researcher and not just by members of the RobotCub project.

In setting out the iCub cognitive architecture, we will proceed in steps. We begin by addressing the phylogenetic capabilities and then address the modulation of these capabilities.

We then move on to consider the issues of prospection and anticipation. This is accomplished in several ways, both using feed-forward control at the level of self-organization in phylogenetic capabilities and by the addition of component in the cognitive architecture to effect rehearsal of possible scenarios — perception/action sequences — but decoupled from the physical sensorimotor control circuits. Finally, we consider how the pivotal requirement of self-modification is satisfied so that the iCub can develop, *i.e.* alter its dynamics over time and as a consequence of its perception/action experiences.

15.6.1 The iCub Phylogeny

The focus of the iCub cognitive architecture is self-development. However, development implies the existence of a basis for development; in other words, ontogenesis requires some initial phylogenetic configuration on which to build. This section presents a non-exhaustive list of initially-planned innate perceptuo-motor and cognitive skills that need to be effected in the iCub in order to facilitate its subsequent development. These skills (or abilities) are based primarily on the results and insights from developmental psychology in Part II and from a walk-through of the empirical investigations derived from the scenarios for development set out in Sections 16 and 17, respectively.

Note that the performance of all phylogenetic abilities may improve with time as their operational parameters adjust with experience. That is, each capability should have some capacity for learning. This adaptivity differs from capabilities that are the result of ontogenesis because there has been little or no modification of the system's state space, *i.e.* they don't arise as a the result of a process of self-modification or development, but by on-line parameter estimation.

The minimal phylogenetic and cognitive capabilities to be developed for the iCub are summarized in Table 4. They are assigned to one of three classes:

1. Those that correspond directly to the scenario capabilities; these will usually be based on a combination of (possibly tuned) phylogenetic capabilities, sub-cortical action-selection capabilities, and cortical prospection capabilities.

2. Those that correspond directly to quasi-independent phylogenetic capabilities.

3. Those that correspond to components of these phylogenetic capabilities; these correspond to re-usable general purpose sensor and control utility functions.

Each capability in the first category will be implemented as a (possibly large) set of communicating YARP executables. Capabilities in the second category may be implemented as a single YARP executable or possibly as a small set of communicating YARP executables. Those from the third category will be implemented as a single general-purpose YARP executable. These should be designed to be re-usable and interoperable since there may be more than one YARP module with the same functionality.

The concept of localization in sensorimotor space is crucial: it implies an ego-centric frame of reference for a location of an entity, calibrated in terms of the motoric — reaching and locomotion — control parameters of the iCub.

Re-orienting pose does not necessarily imply a well-identified control loop; it might simply involve a perturbation of motor space to facilitate reinforcement learning, for instance.

Scenario Capabilities: cognitive perception/action behaviours
Object tracking through occlusion (smooth pursuit & saccades)
Learn to coordinate VOR & tracking
Learn to reach towards a fixation point
Attention and action selection by modulation of capabilities
Conditional modulation based on anticipation
Construct sensorimotor maps & cross-modal maps
Learn by demonstration (crawling & constrained reaching)
Exploratory, curiousity-driven, action
Experience-based action selection based on interaction histories
Navigate based on local landmarks and ego-centric representations
Quasi-independent Phylogenetic Capabilities
Saccadic re-direction of gaze towards salient multi-modal events
Focus attention and direct gaze on human faces
Ocular modulation of head pose to centre eye gaze
Move the hand(s) towards the centre of the visual field
Stabilize & integrate of saccadic percepts
Stabilize gaze with respect to self-motion (VOR)
Create attention-grabbing stimuli
Gait control
Component Capabilities
Compute optical flow
Compute visual motion with ego-motion compensation
Segmentation of the flow-field based on similarity of flow parameters
Segmentation based on the presence of a temporally-persistent boundary
Fixation and vergence
Gaze control: smooth pursuit with prediction; possibly tuned by learning
Classification of groups of entities based on low numbers
Classification of groups of entities based on gross quantity
Detection of mutual gaze
Detection of biological motion

Table 4: Initial phylogenetic and cognitive capabilities.

We represent this collection of innate phylogenetic abilities in the iCub cognitive architecture as a series of arrow circles, in the spirit of Maturana and Varela’s ideogram of a self-organizing (autopoietic) system [MV87]; see Figure 10.

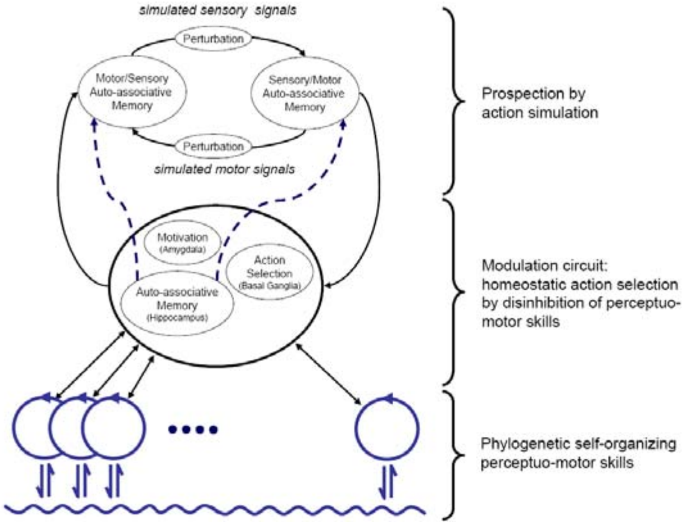

simulated sensory signals
Perturbation
Motor/Sensory Auto-associative Memory
Sensory/Motor Auto-associative Memory
Perturbation
simulated motor signals
Motivation (Amygdala)
Action Selection (Basal Ganglia)
Auto-associative Memory (Hippocampus)
Prospection by action simulation
Modulation circuit: homeostatic action selection by disinhibition of perceptuo-motor skills
Phylogenetic self-organizing perceptuo-motor skills

Figure 10: The iCub cognitive architecture, Version 0.1.

Before proceeding, is worth making a few general remarks about the phylogenetic abilities.

First, all the capabilities should exhibit some form of self-organization by which the sensorimotor contingencies are learned (in the case that no sensorimotor mapping is given *a priori*) or tuned (in the case that some sensorimotor mapping is provided). Self-organization can be viewed as a form of learning. For example, Olsson *et al.* show how it is possible to learn first the information structure of a system's sensors and then the effects that certain settings of actuators have on these sensors [ONP06]. Thus, the *forward model* — the change in sensor data that arises from a change in actuator parameters —is learned. Entropy-based information theoretic methods are used to establish this relationship through an empirical process of motor babbling whereby the system itself autonomously explores its sensorimotor space, learning the association between movements and sensory perceptions.

Second, there may be a need to allow direct interconnection between the distinct phylogenetic perceptuo-motor abilities and enhanced skills without having to revert to the modulation circuit or the prospection circuit. For example, head stabilization with inertial sensing during body motion may require the use of individual feed-forward models which are specific to each of the contributing skills. It may be more appropriate to have some specialized integration of these models rather than depending on the more temporally-extended prospection circuit shown in Figure 10.

Third, our goal is that sensorimotor skills should ideally be modelled as some form of non-linear dynamical system. We can accomplish this using collections of coupled dynamical systems with well-defined attractor properties (*e.g.* attractors, repulsors, limit cycles, *etc.*), typically specified by a system of differential equations, or by identifying a non-linear dynamical equation that captures the macroscopic behaviour of the system. In this case, the behaviour is specified by the system's collective variables (also referred to as order parameters). Ideally, these parameters would be identified in the process of learning sensorimotor contingencies. The collective variables (and the exact form of the dynamical system) define the dynamics of the sensorimotor circuit, its phase space, and its attractor structure. There are also control parameters which represent the environmental perturbations that influence the behaviour the system (but are not constitutents of its dynamical identification and do not specify the system). It is worth remarking that Kelso suggests that in fact a system should be modelled at a minimum of three distinct levels [Kel95]. These are as follows (see also Figure 11).

1. A boundary constraint level that determines the task or goals;
2. A collective variable level that characterizes coordinated states;
3. A component level which forms the realized system.

Kelso argues that the “Boundary constraints, at least in complex biological systems, necessarily mean that the coordination dynamics are context or task dependent”. Take away the context and you take away the basis for the model. Furthermore, the instantiation of the system has a direct role to play in the model itself (which is another way of saying that the system morphology matters and cannot be abstracted away).

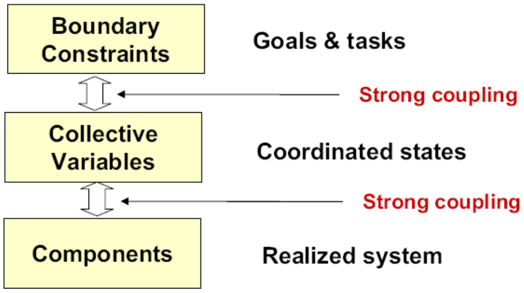


Figure 11: The three levels at which a system should be modelled: a boundary constraint level that determines the task or goal, a collective variable level that characterizes coordinated states, and a component level which forms the realized system (after [Kel95]). All three levels are equally important and should be considered together.

Fourth, the phylogenetic abilities will typically need to exploit both feed-back and feed-forward control. Feed-back control provides very little predictive control (at best, the derivative term in a PD controller can be viewed as a form of predictive control) and exclusive use of feed-back control is inadequate for achieving effective response in biologial systems where the latency in neural processing can be much greater than the time scales required for effective response to external stimuli. Indeed, it has been argued that the brain can be viewed as a way providing effective predictive control [Ber00]. One way of achieving this (but, as we will see, not the only way) is to effect feed-forward control. Feed-forward control is based on a model of the environment, specifically the relationship between collateral variables and the control variables. While feed-back control measures an error in the control variable (or a derivative of it) and responds accordingly, and therefore acts after the control variable has changed, feed-forward control measures the change in a collateral variable that effectively predicts that a change in the control variable will occur. For example, in feed-back cruise control in a car, feed-back control measures the velocity (the control variable), detects an error, and modifies the fuel-injection to reduce the error. A feed-forward controller would instead measure, *e.g.*, the slope of the road, either locally or at some distance in front of the car, and modify the fuel injection *before* there is a change in velocity.

Ideally, the self-organizing sensorimotor phylogenetic capabilities will learn not only the sensorimotor contingencies to effect feed-back control, *e.g.* the collective variables in a dynamical model, but also the collateral factors, *e.g.* the control parameters, that allow feed-forward control.

15.6.2 Modulation of Innate Skills

The perceptuo-motor skills outlined in the previous section operate concurrently, competitively, and cooperatively. A cognitive architecture must specify how these skills are modulated or deployed and how the competition and cooperation is effected.

One plausible approach is suggested by Shanahan [Sha06, SB05, Sha05b, Sha05a] based on global workspace theory [Baa98, Baa02] whereby specialist processes compete and co-operate for access to a global workspace. The winner(s) of the competition gain(s) controlling access to the global access and can then broadcast information back to the competing specialist processes. Shanahan argues that this process allows a sequence of states to emerge from the interaction of many separate parallel processes (see Figure 5).

In the brain, the basal ganglia are responsible for action selection and disinhibition has been proposed as the basic mechanism by which these basal ganglia circuits affect behavior [CVDD85, DC85, HW83]. This suggests that any modulation circuit that is proposed for inclusion in the iCub architecture should take into consideration the function and operation of the basal ganglia, addressing, *e.g.*, reinforcement learning [Doy99], sub-cortical loops with brainstem sensorimotor structures such as the superior colliculus [MSS+05], cortical loops with the neocortex [ADS86], and perhaps some form of short-term memory, possibly effected using an auto-associative structure, for the storage and recall of spatial and episodic events. Rougier, for instance, has proposed and validated an architecture for an auto-associative memory based on the organization of the hippocampus, involving the entorhinal cortex, the dentate gyrus, CA3, and CA1 [Rou01]. A feature of this architecture is that it avoids the catastrophic interference problem normally linked to associative memories through the use of redundancy, orthogonalization, and coarse coding representations. Rougier also notes that the hippocampus plays a role in 'teaching' the neo-cortex, *i.e.* in the formation of neocortical representations. We will return to this point again in Section 15.6.4.

It is noteworthy that the closed loop subcortical and cortical circuit structures are compatible with the global workspace theory for modulation of or selection between competing cognitive, affective, as well as sensorimotor functions.

A basal ganglia model for action selection in a mobile robot is reported in [P*et al.*02].

The question as to what forms the basis for the saliency function which the basal ganglia utilize in making a selection and disinhibiting some sensorimotor circuit remains open. Shanahan suggests the inclusion of the amygdala in the circuit to provide for affective modulation of the action selection process [Sha06].

Figure 10 shows this modulation component of the iCub cognitive architecture with three sub-components: auto-associative memory, action selection, and motivation (reflecting saliency). No interconnections are suggested at this point. What is clear, though, is that the modulation component is connected to each phylogentic skill. These three components are labelled (in parentheses) hippocampus, basal ganglia, and amygdala to denote their biological inspiration. However, we emphasize that is not intended to produce faithful models of these regions.

15.6.3 Prospection and Anticipation

We have emphasized throughout this document that cognition can be viewed as the complement of perception in that it provides a mechanism for choosing effective actions based not on what has happened and is currently happening in the world but based on what may happen at some point in the future. That is, cognition is the mechanism by which the agent achieves an increasingly greater degree of anticipation and prospection as it learns and develops with experience. Although it would be wrong to dismiss perceptual faculties as purely reactive — as we noted in Section 15.6.1 our sensory apparatus provide for some limited predictive capability — some other means is required to anticipate what might happens, especially at longer timescales. One way of achieving this functionality is include a component (or set of circuits) that simulate events and use the outcome of this simulation in guiding actions and action selection. In Berthoz's words 'the brain is a biological simulator that predicts by drawing on memory and making assumptions' … 'perception is simulated action' [Ber00].[11]

This action simulation works concurrently with the innate and learned abilities, and the modulation circuitry, that were described above. In fact, the simulation circuitry provides just another 'input' to this modulation process which can work either competitively or cooperatively with existing skills. Berthoz again:

> The brain processes movement according to two modes. One, conservative, functions continuously like a servo system; the other, projective, stimulates movement by predicting its consequences and choosing the best strategy'.

Another particularly significant feature of this potential capacity for simulation is that it is not structurally coupled with the environment and thereby is not subject to the constraints of real-time interaction that limit the sensori-motor processes [WF86]: the simulation can be effected faster than real-time.

Naturally, the question arises of how one should accomplish —model and implement—this capacity for simulation. Shanahan's work again provides some insights. As noted above, Shanahan's cognitive architecture [Sha06] is comprised of the following components: a first-order sensori-motor loop, closed externally through the world, and a higher-order sensori-motor loop, closed internally through associative memories (see Figure 5). The first-order loop comprises the sensory cortex and the basal ganglia (controlling the motor cortex), together providing a reactive action-selection sub-system. The second-order loop comprises two associative cortex elements which carry out off-line simulations of the system's sensory and motor behaviour, respectively. The first associative cortex simulates a motor output while the second simulates the sensory stimulus expected to follow from a given motor output. The higher-order loop effectively modulates basal ganglia action selection in the first-order loop via an affect-driven amygdala component. Thus, this cognitive architecture is able to anticipate and plan for potential behaviour through its associative internal sensori-motor simulation.

Figure 10 shows the prospective action simulation component of the iCub cognitive architecture with two sub-components in the same vein as Shanahan: a sensory hertero-associative memory that receives efferent (motor) input produces afferent (sensory) output. This feeds into a motor heteroassociative memory that in turn produces (simulated) efferent (motor) output. This output is connected recurrently back to the sensory associative memory and also back to the modulation circuit.

Since some form of action selection mechanism is also required in this circuit, just as it is in the primary modulation circuit, two unspecified perturbation components have been added to the interface between the two associative memories. This also allows for some element of innovation in the perception and action signals.

The prospection circuit also implies some capacity for recognition, inference, and communication. This is implicit at present and requires further analysis.

It may be worth remarking here on the difference between auto-associative memory and heteroassociative memory (often written simply as associative memory, without the qualification). An auto-associative memory takes as input a vector of data and by association produces a different, typically more complete, version of the same vector; that is, it performs pattern completion. A heteroassociative memory in contrast takes as input a vector of data and produces by association a different vector of data; typically the two vectors correspond to different vector spaces (*e.g.* the space of afferent

[11] Berthoz's statement [Ber00] that 'perception is simulated action' is reminiscent of Max Clowes's assertion that 'perception is controlled hallucination'.

sensory data and the space of efferent motor data).

15.6.4 Self-Modification

We come finally to a crucial aspect of developmental emergent cognition: ability to self-modify. There are two aspects to this: the mechanism of self-modification and the basis (or drive) for the self-modification.

Learning is tightly tied up with mechanisms for self-modification. Three types of learning can be distinguished: supervised learning in which the teaching signals are directional error signals, reinforcement learning in which the teaching signals are scalar rewards or reinforcement signals, and unsupervised learning with no teaching signals. Doya argues that the cerebellum is specialized for supervised learning, basal ganglia for reinforcement learning, and the cerebral cortex for unsupervised learning [Doy99]. He suggests that in developing (cognitive) architectures, the supervised learning modules in the cerebellum can be used as an internal model of the environment and as short-cut models of input-output mappings that have been acquired elsewhere in the brain. Reinforcement learning modules in the basal ganglia are used to evaluate a given state and thereby to select an action. The unsupervised modules in the cerebral cortex represent the state of the external environment as well as internal context, providing also a common representational framework for the cerebellum and the basal ganglia which have no direct anatomical connections.

We need to distinguish carefully between learning in the sense of adjusting or improving innate or existing skills, and learning in the sense of adjusting the systems structure, organization, or operation with a view to accommodating new skills and actions. Both are required in a cognitive system but will recruit different mechanisms and will be driven by different criteria. We have already alluded to the former type of learning in Section 15.6.1 under the heading of *the* iCub *phylogeny*. Learning can be effected as part of the self-organizational process inherent in each innate skill, perhaps effected by supervised learning in the manner of the cerebellum. We speculate that the enhanced phylogenentic skills are learned through reinforcement learning in the modulation circuitry outlined in Section 15.6.2, specifically by hippocampus auto-associative memory and basal ganglia action selection mechanism.

This leaves us with the learning associated with development and self-modification. Before suggesting a mechanism, we turn first to the matter of what drives the process of self-modification. We noted in Section 15.5 that development should be driven by both exploratory 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. There remains the problem though of exactly *how* we can measure any advancement of the system's understanding of novel regularities, of the system's actions, and interaction. That is, we require a metric that allows one to drive the development in the right direction, even though the learning involved in development is likely to be non-monotonic (*i.e.* it will often exhibit short-term failure before long-term success).

We speculate here that such a metric might be founded on the same principles as that used for learning sensorimotor contingencies, *i.e.* an entropy-based metric and specifically a normalized entropy reduction metric which indicates that the system's perceptuo-motors space — both actual and simulated — is more ordered, even when normalized by some function of its increased space of potential actions.

This in turn suggests a (highly-speculative) mechanism for self-development. It is plausible that the experience gathered by reinforcement learning and encapsulated in the auto-associative memory of the modulation circuits — a process that involves not only modulation of the phylogentic and enhanced phylogentic skills but also the inputs from the prospective hetero-associative circuits — may periodically update the long-term hetero-associative memories, thereby giving rise to an increased space of potential (simulated prospective) action, which in turn drives the system's actions and experiences further, increasing its effectiveness and resilience. This memory-memory update would be modulated by on the basis of the entropy-reduction metric. Since such an update process should not interfere with the normal operation of the cognitive system, we *speculate* that this update happens when the system is in a rest state, *i.e.* when it is sleeping. Figure 10 indicates this developmental process by showing blue return arrows from the modulation circuits to the prospection circuits. McClelland *et al.* have suggested a similar process. They note that the hippocampal formation and the neo-cortex form a complementary system for learning [MNO95]. The hippocampus facilitates rapid auto- and heteroassociative learning which is used to reinstate and consolidate learned memories in the neo-cortex in a gradual manner. In this way, the hippocampal memory can be viewed not just as a memory store but as a 'teacher of the neo-cortical processing system'. Note also that the reinstatement can occur on-line, thereby enabling the overt control of behavioural responses, as well as off-line in, *e.g.* active rehearsal, reminiscence, and sleep.

Before closing this section, it is worth remarking that although we have drawn heavily in creating the iCub cognitive architecture on work by Shanahan, it is significant that his own cognitive architecture does not (yet) incorporate any learning mechanisms.

15.6.5 Candidate Cognitive Mechanisms

With this general architecture established, we need to be more explict about the actual mechanisms that will be used to effect the action selection, sensorimotor fusion, and sensorimotor simulation. To this end, we have developed the architecture somewhat as shown in Figure 12. The candidate mechanisms identified in this figure are described in [VSM07].

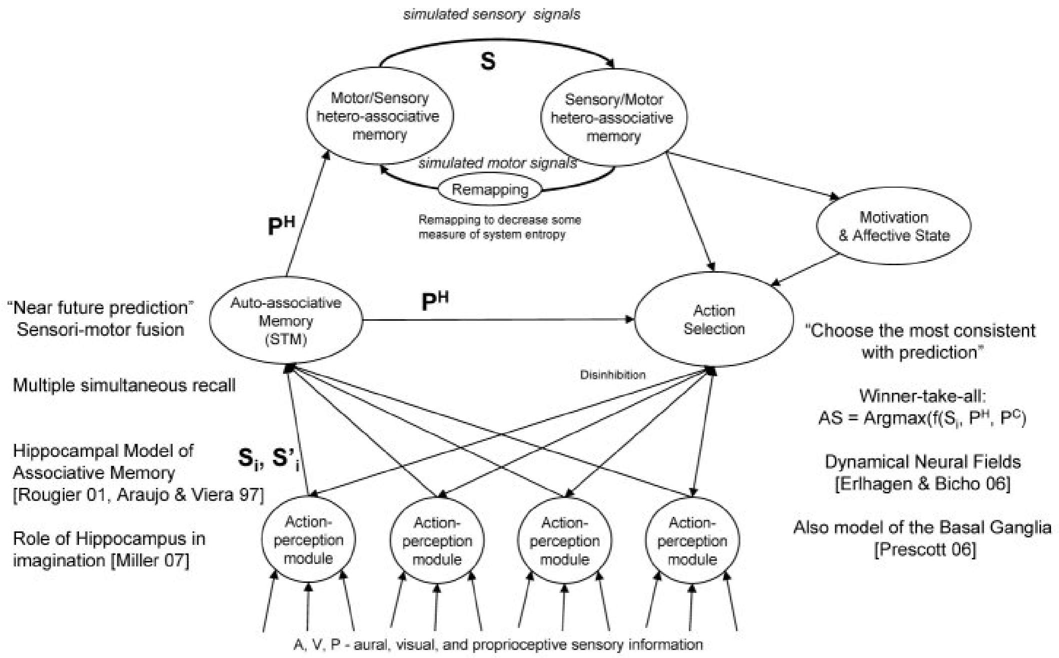

simulated sensory signals
S
Motor/Sensory hetero-associative memory
Sensory/Motor hetero-associative memory
simulated motor signals
Remapping
Remapping to decrease some measure of system entropy
Motivation & Affective State
P^H
"Near future prediction" Sensori-motor fusion
Auto-associative Memory (STM)
P^H
Action Selection
"Choose the most consistent with prediction"
Disinhibition
Multiple simultaneous recall
Winner-take-all: AS = Argmax(f(S_i, P^H, P^C)
Hippocampal Model of Associative Memory [Rougier 01, Araujo & Viera 97]
S_i, S'_i
Dynamical Neural Fields [Erlhagen & Bicho 06]
Action-perception module
Action-perception module
Action-perception module
Action-perception module
Role of Hippocampus in imagination [Miller 07]
Also model of the Basal Ganglia [Prescott 06]
A, V, P - aural, visual, and proprioceptive sensory information

Figure 12: The iCub cognitive architecture, Version 0.2.

15.6.6 Realization of an Essential Phylogeny

While the foregoing cognitive architecture requirements and initial design were being pursued, there was also a need to develop an initial architecture that will allow the action-perception modules shown in Figure 12 to be integrated in a way that makes sense from a phylgenetic perspective and that is meaningful for both neuroscience and developmental psychology. In other words, there was a need to build a minimal functioning system. This gave rise to a parallel design exercise, with one strand focussing the principles governing the design of the cognitive architecture and with the other strand focussing on the implementation of an essential core of functionality. This second strand – referred to as the software architecture – provided a concrete way of grounding the cognitive architecture design and it was driven by the empirical investigations that we have formulated to investigate specific phylogenetic skills and ontogenetic development processes associated with a number of developmental scenarios. These scenarios are described in Section 16 while the empirical investigations are described in Section 17. Consequentially, this software architecture allowed us to work on building a software system which is neuro-scientifically and psychologically plausible, biased towards the very early phylogenetically-derived behaviours but nevertheless supportive of subsequent developmental ontogeny.

Over a period of two years, this software architecture underwent four revisions. The final revision – Version 0.4 – is shown in Figure 13; previous versions can be found on the iCub wiki at http://eris.liralab.it/wiki/ICub_Cognitive_Architecture#Links.

This software architecture is essentially a salience-based 7 degree-of-freedom gaze-controlled reaching system. Most of the work that went into the four revisions was concerned with ensuring that the forward and reverse connections between various components were consistent with what is known about the neurophysiology and psychology of the brain. The cognitive components are essentially placeholders (denoted semantic modulation in Figure 13).

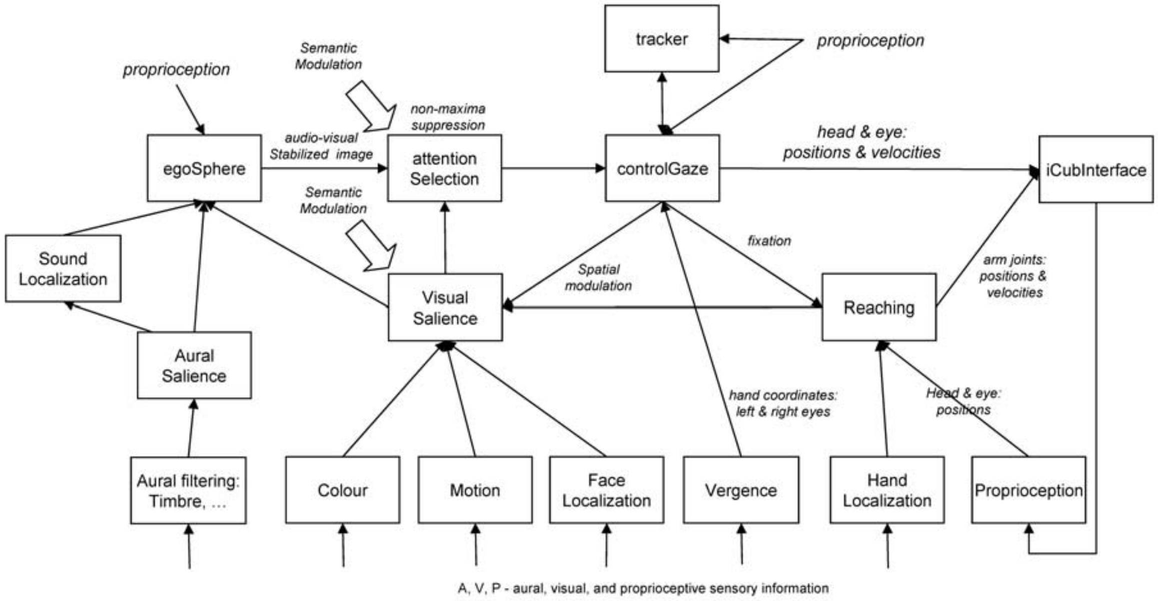


Figure 13: The iCub software architecture, Version 0.4.

15.6.7 Realization of the iCub Cognitive Architecture

> "Il faut reculer pour mieux sauter"
> [One must draw back in order to make a better attack]
> Michel Eyquem de Montaigne, *Essays* (Bk. I, ch. XXXVIII)

Enactive Cognition

At this point, we have established an outline design for a cognitive architecture, soundly based on the principles of neuro-physiology and developmental psychology, and a more detailed design embracing the essential phylogeny, the software architecture. The work during 2009 was devoted to a concerted effort to drive the convergence of the software architecture and the cognitive architecture, the subsequent specification of a revised cognitive architecture, followed by the realization of this architecture as a collection of YARP iCub modules. To achieve this without compromising on the project's commitment to neuro-physiologically plausible realization and a philosophical commitment to action-dependent embodied cognition, i.e. enaction, it was necessary to stand back a little to highlight the core essence of the iCub approach to cognition. We will briefly review these essentials in the next few paragraphs before proceeding to discuss the changes that were subsequently made to these initial cognitive and software architectures to realize the final iCub cognitive architecture.

Enactive systems are based on five central principles: embodiment, experience, emergence, autonomy, and sense-making [VMS09]. Cognition is the process by which the issues that are important for the continued operation of a cognitive entity are brought out or enacted: co-determined by the entity as it interacts with the environment in which it is embedded. An enactive cognitive agent is embodied and embedded in the environment and is specified by it, while, at the same time, the process of cognition determines what is real or meaningful for the agent. Ultimately, this means that the system's perceptions reflect the actions which are consistent with the maintenance of the system's autonomy. Thus, an enactive cognitive agent constructs its reality as a result of its operation in that world and therefore cognitive understanding is intrinsically specific to the embodiment of the system and dependent on the system's history of interactions, i.e., its experiences. Thus, nothing is 'pre-given'. Instead there is an enactive interpretation: a real-time context-based choosing of relevance. This is often referred to as 'sense-making'. For enactive systems, the purpose of cognition is to uncover unspecified regularity and order that can then be construed as meaningful because they facilitate the continuing operation, development, and autonomy of the cognitive system.

For an enactive system, knowledge is the effective use of sensorimotor contingencies grounded in the structural coupling of the system with its environment. Knowledge is particular to the system's history of interaction. If that knowledge is shared among a society of cognitive agents, it is not because of any intrinsic abstract universality, but because of the consensual history of experiences shared between cognitive agents with similar phylogeny and compatible ontogeny. The knowledge possessed by an enactive system is built on sensorimotor associations, achieved initially by exploration, and affordances.

Internal Simulation

An enactive system uses the knowledge gained to form new knowledge which is then subjected to empirical validation to see whether or not it is warranted (we, as enactive beings, imagine many things but not everything we imagine is plausible or corresponds well with reality, i.e. our phenomenological experience of our environment). One of the key issues in cognition, in general, and enaction, in particular, is the importance of *internal simulation* in accelerating the scaffolding of this early developmentally-acquired sensorimotor knowledge to provide a means to *predict* future events, *reconstruct* (or explain) observed events (constructing a causal chain leading to that event), and *imagine* new events. Crucially, there is a need to focus on re-grounding predicted, reconstructed, or imagined events in experience so that the system — the robot — can do something new and interact with the environment in a new way.

This reappraisal re-focussed our efforts in modeling cognition, in general, and internal simulation, in particular, to provide capabilities for prediction, reconstruction, and imagination. In addition, cognitive motivation encapsulated in the system's affective state were made more explicit so that they address curiosity (dominated by exogenous factors), exploration (dominated by endogenous factors), and social engagement (where exogenous and endogenous factors balance). This distinction between the exogenous and the endogenous highlighted the need to modify the attention system to incorporate both factors.

These considerations led to significant changes in the cognitive architecture and the rationalization of the software architecture and the cognitive architecture. This rationalization occurred progressively throughout the year, with changes being consolidated at three project meetings in Lisbon (April 2009), Sestri Levanti (July 2009), and Genoa (November 2009). These developments and the current version of the cognitive architecture (v. 0.4) are documented fully on the iCub wiki.[12] The rationalization itself involved the adaptation and extension of the old software architecture which encapsulated gaze, reaching, and locomotion capabilities into a more comprehensive architecture that incorporated the key components of the original cognitive architecture. This revised cognitive architecture thus became the first iteration of a blueprint for the realization of the cognitive architecture as a set of YARP iCub modules. The term software architecture then reverted to its original meaning as the YARP middleware system.

A major change the revised cognitive architecture involved the replacement of the internal simulation area which had been inspired by Shanahan's coupled hetero-associative memories with a new approach based on auto-associative perceptual memory and hetero-associative event memory. These were subsequently re-cast as an *episodic memory* and a *procedural memory* following the Sestri Levante meeting in July 2009.

Episodic Memory

The episodic memory is a simple memory of autobiographical events. It is a form on one-shot learning and does not generalize multiple instances of an observed event. That functionality will be provided later by some form of semantic memory. In its current form, the episodic memory is unimodal (visual). In the future, as we develop the iCub cognitive architecture, it will embrace other modalities such as sound and haptic sensing. It will also include some memory of emotion. This fully-fledged episodic memory will probably comprise a collection of unimodal auto-associative memories connected by a hetero-associative network (see Section 15.6.10 below). The current version implements a simple form of content-addressable memory based on colour histograms and log-polar mapping. The motivation for these choices is as follows.

In many circumstances, it is necessary to have an iconic memory of landmark appearance that is scale, rotation, and translation invariant (SRT-invariant) so that landmarks can be recognized from any distance or viewing angle. Depending on the application, a landmark can be considered to be an object or salient appearance-based feature in the scene. For our purposes with the iCub cognitive architecture, translation invariance — which would facilitate landmark recognition at any position in the image — is not required if the camera gaze is always directed towards the landmark. This is the case here because gaze is controlled independently by a salience-based visual attention system. There are three components of rotation invariance, one about each axis. Rotation about the principal axis of the camera (i.e. roll) is important as the iCub head can tilt from side to side. Rotation about the other two axes reflects different viewpoints (or object rotation, if the focus of attention is an object). Typically, for landmarks, invariance to these two remaining rotations is less significant here as the orientation of objects or landmarks won't change significantly during a given task. Of course, full rotation invariance would be best. Scale

[12] http://eris.liralab.it/wiki/ICub_Cognitive_Architecture

invariance, however, is critical because the apparent size of the landmark patterns may vary significantly with distance due to the projective nature of the imaging system. There are many possibilities for SRT-invariant representations but we have used colour histograms as the invariant landmark representation and matching will be effected using colour histograms and (a variant of) colour histogram intersection, respectively [SB90, SB91]. Colour histograms are scale invariant, translation invariant, and invariant to rotation about the principal axis of the camera (i.e. the gaze direction). They are also relatively robust to slight rotations about the remaining two axes. Colour histogram representation and matching strategy also have the advantage of being robust to occlusion. They are also robust to variations in lighting conditions, provided an appropriate colour space is used. We use the HSV colour space and use the H and S components only in the histogram.

The episodic memory operates as follows. When an image is presented to the memory, if a previously-stored image matches the presented image sufficiently well (based on the Bhattacharyya distance metric), the stored image is recalled; otherwise, the presented image is stored. The images presented to the module can be either conventional Cartesian images or Log-polar mapped images. We use log-polar images in the iCub cognitive architecture as they are effectively centre-weighted due to the non-linear sampling and low-pass filtered at the periphery. This makes it possible to effect appearance-based image/object recognition without prior segmentation.

Procedural Memory

The procedural memory is a network of associations between action events and pairs of perception events. For the moment, a perception event is a visual landmark which has been learned by the iCub and stored in the episodic memory. An action event is a gaze saccade with an optional reaching movement, a hand-pushing movement, a grasping movement, or a locomotion movement. Since the episodic memory effects one-shot learning, it has no capacity for generalization. This generalization will be effected at some future point by the long-term 'semantic' memory and it may be appropriate then to link the procedural memory to the long-term memory. This will be particularly relevant in instances where the procedural memory is used to learn affordances. A clique in this network of associations represents some perception-action sequence. This clique might be a perception-action tuple, a perception-action-perception triple, or a more extended perception-action sequence. Thus, the procedural memory encapsulates a set of learned temporal behaviours (or sensorimotor skills, if you prefer). The procedural memory can be considered to a form of extended hetero-associative memory (hetero because the recalled information or vector is not necessarily in the same space as the information used to effect the recall).

The procedural memory has three modes of operation, one concerned with learning and two concerned with recall. In the learning mode, the memory learns to associate a temporally-ordered pair of images (perceptions) and the action that led from the first image perception to the second. In recall mode, the memory is presented with just one image perception and an associated perception-action-perception (P_i, A_j, P_k) triple is recalled. There are two possibilities in the mode: (1) The image perception presented to the memory represents the first perception in the (P_i, A_j, P_k) triple; in this case the recalled triple is a prediction of the next perception and the associated action leading to it. (2) The image perception presented to the memory represents the second perception in the (P_i, A_j, P_k) triple; in this case the recalled triple is a reconstruction that recalls a perception and an action that could have led to the presented perception. In both prediction and reconstruction recall modes, the procedural memory produces as output a (P_i, A_j, P_k) triple, effectively completing the missing tuples or (P_i, ~, ~) or (~,~, P_k).

Further Modifications

Other changes to the software architecture as it merged with the cognitive architecture include the removal of the following components:

- tracker (to be handled instead by attention/salience sub-system)
- face localization (to be handled instead by attention/salience sub-system)
- hand localization (to be handled instead by attention/salience sub-system)
- sound localization (to be handled by salience module)

and the addition of the following components

- Exogenous Salience
- Endogenous Salience
- Locomotion
- Matching
- Auto-associative memory episodic memory
- Hetero-associative procedural memory
- Affective state
- Action selection

This rationalization of version 0.2 of the cognitive architecture (Figure 12) with version 0.4 of the software architecture (Figure 13) led initially to version 0.3 of the cognitive architectures (Figure 14) and ultimately to version 0.4 of the cognitive architecture (Figure 15).

In the next section, we will address the software implementation of this cognitive architecture and the specification of each component as an individual YARP iCub module.

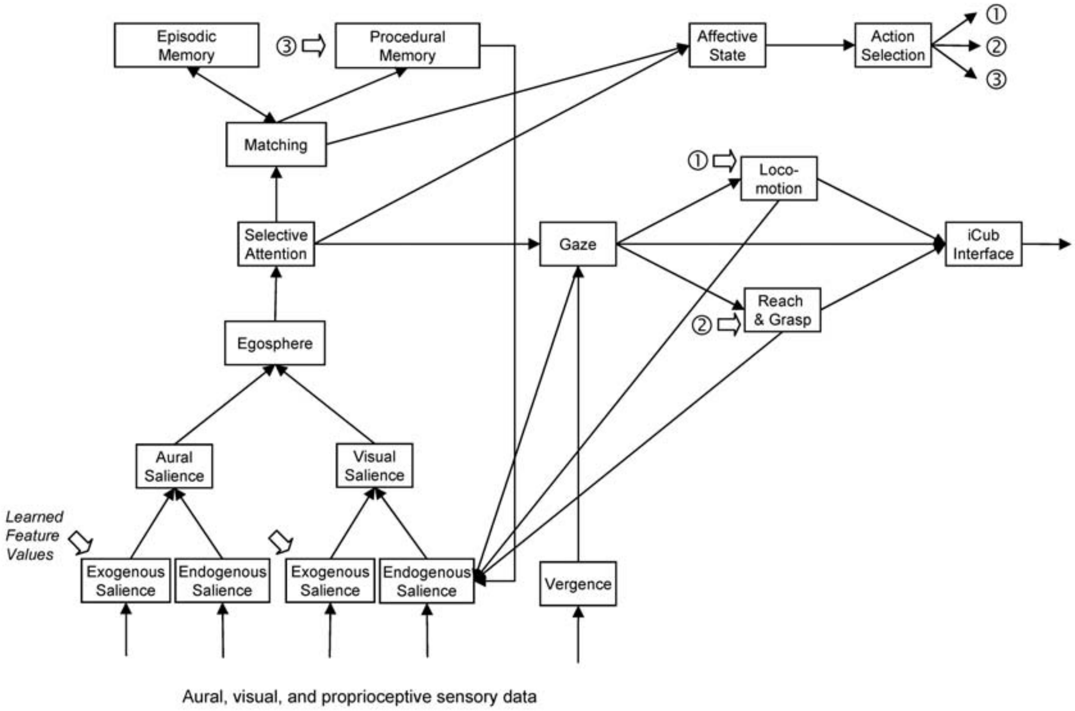


Figure 14: The iCub cognitive architecture, Version 0.3.

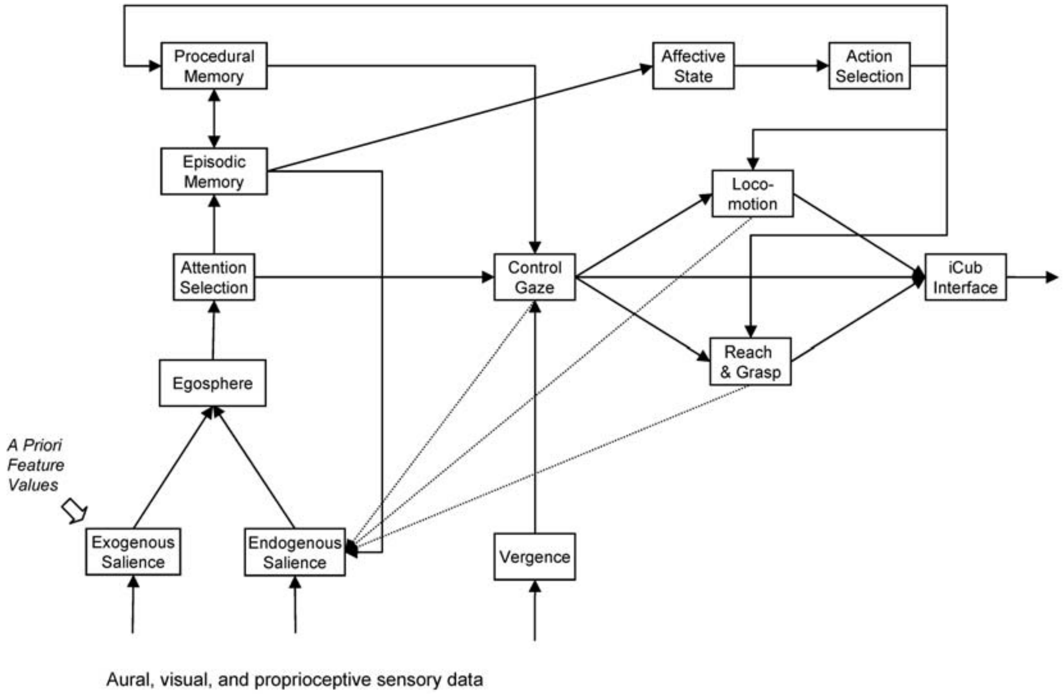


Figure 15: The iCub cognitive architecture, Version 0.4.

15.6.8 Implementation of the Cognitive Architecture

Beginning with the VVV 09 Summer School in Sestri Levante in July 2009, we undertook a substantial effort to realize the revised cognitive architecture (initially version 0.3 as shown in Figure 14 and subsequently in version 0.4 as shown in Figure 15) as a complete software system comprising an integrated collection of YARP iCub modules.

These modules comprise the following.

- *salience*
- *endogenousSalience (work in progress)*
- *egoSphere*
- *attentionSelection*
- *controlGaze2*
- *episodicMemory*
- *proceduralMemory*
- *crossPowerSpectrumVergence*
- *actionSelection (work in progress)*
- *affectiveState (work in progress)*

The *salience, egoSphere, attentionSelection,* and *controlGaze2* modules were developed at IST, Lisbon, and were a key factor in the realization of the initial software architecture (version 0.4 and previous versions).

The remaining modules were developed at UGDIST and IIT, Genoa, during and after the VVV '09 Summer School in Sestri Levante.

In addition, several support modules were developed as part of this implementation effort. These comprise the following.

- *cameraCalib*
- *rectification*
- *logPolarTransform*
- *imageSource*
- *autoAssociativeMemory*
- *myModule*

These modules are integrated as a single iCub application – *cognitiveGaze* – which is shown in Figures 16 and 17.

A specification of each of the main modules may be found on the iCub wiki at http://eris.liralab.it/wiki/ICub_Cognitive_Architecture.

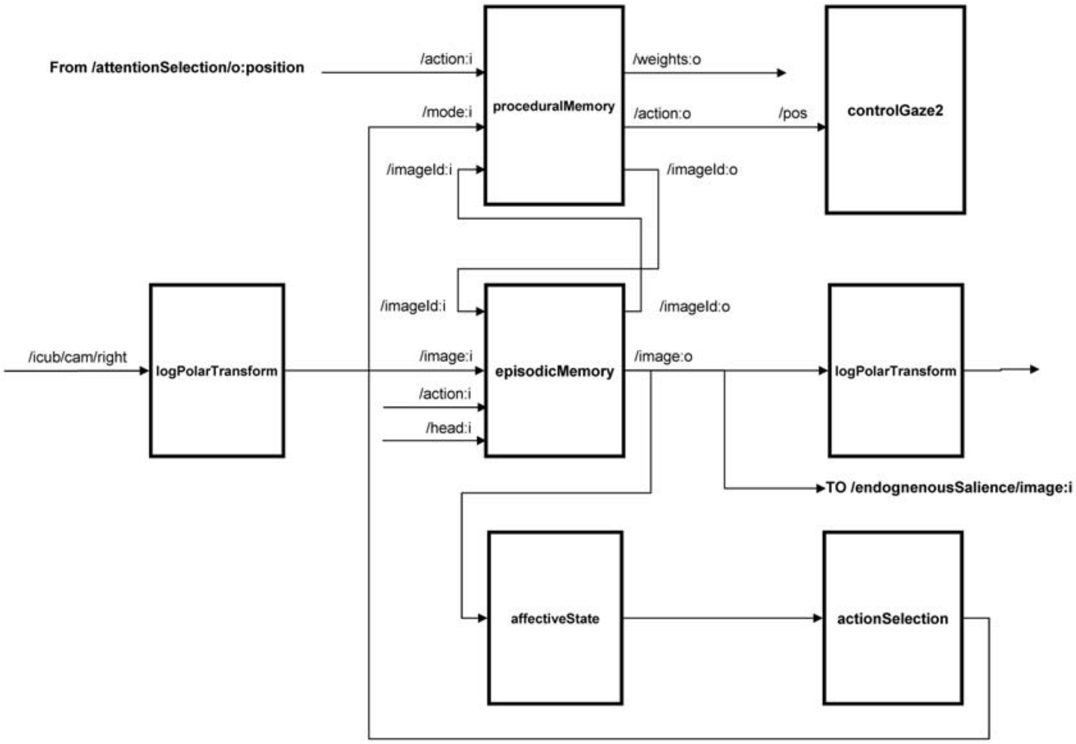


Figure 16: Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part A)

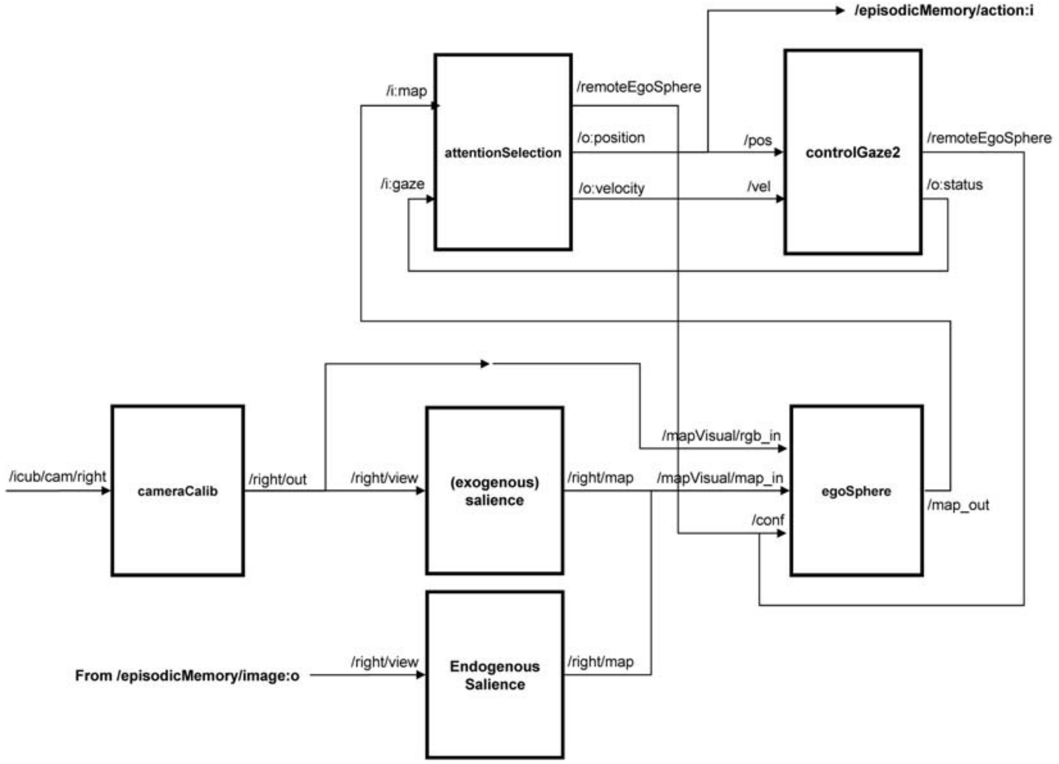


Figure 17: Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part B)

15.6.9 The iCub Cognitive Architecture and the Posner Test

Given that the iCub cognitive architecture has its roots in neurophysiology and developmental psychology, we decided to adapt a standard test to evaluate the operation of the architecture. In particular, we decided to use the Attention Network Test (ANT) developed by Michael Posner and his co-workers [FMS+02]. This test is designed to assess the three attentional networks associated with the functions of alerting, orienting, and executive control. The test was designed to obtain a measure of the efficiency of each of the networks and to be simple enough to be used with children, patients, and animals. The alerting function is defined as achieving and maintaining an alert state, orienting is the selection of information from sensory input, and executive control is defined as resolving conflict among responses [FMS+02].

The ANT requires a participant to determine whether a central arrow points to the left or to the right when displayed on a monitor (see Fig. 18 (b)) and press a corresponding key on a keyboard. The arrow is set among flankers, two to the left, and two to the right. These flankers may be straight lines (neutral), arrows of the same orientation (congruent), or arrows of the opposite direction (incongruent). The display of the arrow is preceded first by the display of a fixation point in the centre of the screen, second by the display of a one of four cue conditions, and third by the fixation point again (see Fig. 18 (c)). After the orientation of the arrow has been determined, or a time limit of 1700 ms has been reached, the fixation point is displayed once again. The four cue conditions are no cue, centre cue, double cue, and spatial cue at the position where the arrow will appear (see Fig. 18 (a)).

Before taking the test, participants are allowed a practice session. Results with adults show that the reaction time is approximately the same for the neutral and congruent target condition, and significantly greater for the incongruent condition. Additionally, the reaction time varies consistently for all three cases, depending on the cue condition, with no cue having the longest reaction time, followed by centre cue, double cue, and with the spatial cue having the shortest reaction time. Error rates are low for the neutral case (approx. 1.25%), slightly lower for the congruent case (approx. 1%), and significantly greater for the incongruent case (approx.4%).

A variant of the ANT for use with children uses animated images of fish rather than static arrows. The reaction time with children is much longer but the pattern of reaction times as a function of flanker and cue is consistent with the adult case.

Our goal is to adapt this test for use with the iCub by using coloured tokens replacing the arrowheads so that we can distinguish them on the basis of hue rather than shape. At time of writing, no results are yet available.

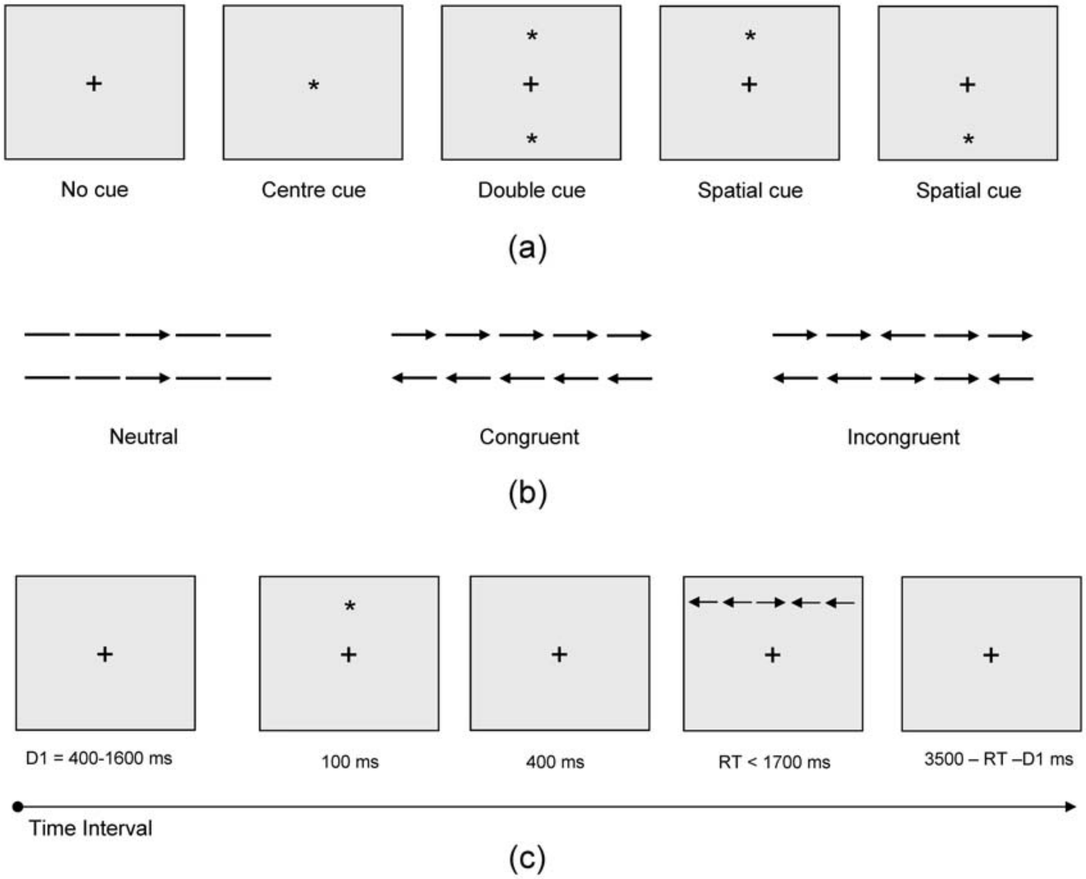


Figure 18: The Posner Test: (a) the four cue conditions; (b) the six stimuli; (c) the timing of the presentation of the fixation point, cue, fixation point, stimulus, and fixation point (adapted from [FMS+02]).

15.6.10 Future Work

Prediction, Reconstruction, and Action: Learning Affordances

Every action entails a prediction about how the perceptual world will change as a consequence of that action. Equivalently, every pair of perceptions is intrinsically linked or associated with an action. So, if we think of a perception-action-perception triplet of associations (P_i, A, P_j), we can effect prediction, reconstruction (or explanation), and action as associative recall by presenting (P_i, A, ~), (~, A, P_j), or (P_i, ~, P_j), respectively, to the procedural memory. In principle, this triplet-based representation is very similar to the iCub framework for learning object affordances (see Deliverable 4.1, pp. 16-20). Here, affordances are represented by a triplet (O, A, E), where O is an object, A is an action performed on that object, and E is the effect of that action. (O, A) → E is the predictive aspect of affordance; (O, E) → A recognizes an action and aids planning; (A, E) → O is object recognition and selection. In the future, we will investigate how this affordance work can be integrated with the cognitive architecture, in general, and the procedural memory, in particular.

Scan-path based Object Representation

There is no explicit concept of objecthood in the iCub cognitive architecture. Arguably, however, parts of a visual scene assume objecthood when they present a persistent and stable pattern of salience. This stable pattern of salience can be encapsulated by a repeatable localized eye gaze scan path pattern and represented by a given (P_a, A_i, P_b … A_j, P_c) clique within the network of associations in the procedural memory. Object recognition then becomes a matter of associative clique retrieval based one all or part of the clique. Again, this is something that will be investigated in the near future.

Locomotion

For locomotion, the procedural memory produces a series of scale-invariant landmarks that should be followed to take the robot from an initial position to a final goal position. These can be learned as the iCub moves about the environment, storing landmarks in its episodic memory as it goes. Since the procedural memory assumes the same image landmark representation as the episodic memory, it simply stores the episodic memory identification number. In the case of locomotion, the procedural memory action events connote the visibility of one landmark from another, and thus connote whether or not one can move directly from one landmark to another. The initial position is input as a scale-invariant landmark image from the attention module: this will typically be the target object to which the robot has navigated. The final position is input also as a scale-invariant landmark representation from the episodic memory: this will typically be the first landmark the robot encountered on its exploratory journey in search of the target. The procedural memory will then produce a sequence of events (landmark images and movements) that the robot should follow to achieve its goal path. There are two options open in generating this procedural sequence. The first is to retrace the landmarks encountered when searching for the target, in the reverse order in which they were encountered. The second is to determine a shortest path between the initial and final positions. The first approach requires no cognitive ability: it is simply a pair-wise association between landmarks. The second approach can be argued to offer a simple cognitive capability by prospectively seeking an optimal set of associations between landmarks, minimizing some overall cost of returning to the goal position.

Action Representation

So far, we have assumed that the action events that are stored in the procedural memory are relative gaze saccades with tags to denote movements (reaching, grasp & object contact, locomotion, or no movement). Recall that iCub cognition involves an implicit model of motion control, specifically the so-called motor-motor control model whereby the proprioceptive state of one set of motors implicitly defines and controls the state of another set during action. For example, a reaching or a locomotive action is specified by the gaze of the eyes: you reach where you are looking or you move to where you are looking. Consequently, it isn't necessary to store the detailed kinematics or dynamics of either locomotion or reaching actions in the procedural memory. Instead, it is sufficient to store simply a tag denoting the type of action. Gaze actions capture the spatial relationships between percepts and, together with the movement tags, specify the actions that perturb the environment, i.e. grasp and object contact motions, are typically object-specific and can be considered to be a form of proprioceptive image of the interaction.

Prediction, Reconstruction, and Imagination: Self-Development

At present, all associations are based purely on the interaction history of the iCub. This needs to be augmented with a process whereby associations can be formed internally by the iCub to facilitate the 'imagination' aspect of cognition. One possible approach is to implement some form of Hebbian learning whereby events that loosely co-occur (i.e. that fire closely in time but are not causally connected) might be associated. Another approach would be to allow the procedural memory to self-modify – i.e. alter the association weights – by establishing cliques within the network of associations that exhibit some form of order, e.g. though an entropy-reduction process, similar to the process of bisociation.

Recursive Events

Allow some form of recursive definition of an event, so that an event itself could be some network of perception-action associations, and not just either an atomic perception or an action as it is at the moment, i.e., generalize the input to the memory to allow something more flexible that the current (P_i, A, P_j) triplet.

Multi-modal Episodic Memory

As currently specified, this auto-associative memory is fairly simple and there are a few natural ways in which it could be extended or augmented. One obvious requirement, especially in the context of the cognitive architecture attention sub-system, is the need to include aural information. One way to do this would be to extend the auto-associative memory to be a multi-modal auto-associative memory, with a composite audio-visual storage and recall. This has the disadvantage of necessarily associating sound and vision with every data set, even though no significant sound may be present for that image (and vice versa). An alternative would be to implement an explicit aural auto-associative memory and link them with a hetero-associative memory.

Generalization

The episodic memory might be extended by implementing some form of generalization. At present, the memory simply does one-shot learning and similar images (or images of similar data) are not generalized. Such one-shot learning based memory is sometimes referred to as episodic memory while memory that consolidates multiple experiences of the same memory is often referred to as semantic memory. Together, they form (according to some psychologists) a form of explicit declarative memory. This is in contradistinction to implicit procedural memory which encapsulates temporal sequencing and skill-based learning. The question this is whether this ability to generalize should be encapsulated or subsumed into the auto-associative memory. Neuroscientific evidence suggests not. For example, McClelland et al. have suggested that the hippocampal formation and the neocortex form a complementary system for learning [MNO95]. The hippocampus facilitates rapid autoassociative and heteroassociative learning which is used to reinstate and consolidate learned memories in the neocortex in a gradual manner. In this way, the hippocampal memory can be viewed not just as a memory store but as a "teacher of the neocortical processing system." This suggests that the best way to proceed would be to implement a separate long-term semantic/generalized memory which takes as input the output of the current episodic memory.

Affective State and Action Selection

Affective state should influence more than just action selection and should be an aspect of the episodic memory so that emotions are associated with events. Once affective state is incorporated into the auto-associated episodic memory, and by extension in the procedural memory, affective state will implicitly modulate salience so that certain features can be modulated by motivations

16 The iCub Ontogeny: Scenarios for Development

A new set of scenarios has been developed to form the basis of the ontogenesis of the iCub . They modify and extend those that were envisaged when the RobotCub technical annex was originally written. The scenarios are enacted or put into practice in the empirical investigations that are detailed in Section 17.

The primary focus of the early stages of ontogenesis is to develop manipulative action based on visuomotor mapping, learning to decouple motor synergies (*e.g.* grasping and reaching), anticipation of goal states, learning affordances, interaction with other agents through social motives, and imitative learning. Needless to say, ontogenesis and development are progressive. In the following, we emphasize the early phases of development, building on the enhanced phylogenetic skills outlined in the Section 15.6.1 and scaffolding the cognitive abilities of the iCub to achieve greater prospection and increased (action-dependent) understanding of the iCub of its environment and mutual understanding with other cognitive agents.

It is important to emphasize that the development program that we intend to use to facilitate the ontongenesis of the iCub is biologically inspired and tries to be as faithful as possible to the ontogenesis of neonates. Consequently, the development of manipulative action will build primarily on visual-motor mapping. The following are the scenarios that will be used to provide opportunities for the iCub to develop, in order of their deployment over time.

Reaching for Objects

The most basic skill is not to grasp the object but to get the hand to the object. In order to do that, the visual system has to define the position of the object in front of it in motor terms. The newborn infant has such an ability. Newborns can monitor the position of the hand in front of them and guide it towards the position of an object. The visual guidance of the hand is crude to begin with and it needs to be trained. Putting the hand into the visual field opens up a window for such learning. When newborn infants approach an object, all the extensors of the arm and hand move in extension synergy. In order to grasp the object, the infant has to overcome this synergy and flex the fingers around the object when the arm is in an extended position. Note that human infants do not master this decoupling of extension and flexion until 4 months of age.

Grasping Objects

Once the iCub masters the extension of the hand towards objects in the surrounding and can flex the fingers around them, grasping skills can develop. However, the iCub must have some kind of motive for grasping objects in order to make this happen. Note that it is the sight of the object that should elicit anticipations of the sensory consequences of the action. Infants who are at the transition to mastering the grasping of objects anticipate crudely the required orientation of the hand. They open the hand fully when approaching any object which optimizes the chances of getting the object into the hand. Adjusting the opening of the hand during the approach to the size of the object to be grasped develops as the infant becomes experienced with object manipulation. The timing of the grasp is controlled visually but, to begin with, at the expense of interrupting the flow of the action (the movement is temporarily stopped before the close around it). This coordination also improves as a function of experience.

Affordance-based Grasping

Grasping objects as a function of their use only develops after infants master reaching and grasping objects in a versatile way (towards the end of the first year of life). The first manipulative actions are general and explorative: squeezing, turning, shaking, putting into the other hand *etc.* The purpose can be said to learn about object properties. More specific and advanced object manipulation skills only develop after the end of the first year of life, like putting objects into apertures, inserting one object into another, position lids on pans, building towers of blocks.

Mastering actions like that relies on anticipation of goal states of manipulatory actions. This is how we intend the iCub to develop its manipulatory action. The sensory effects of manipulatory action should be primarily visual, like the disappearance of the object into the hole.

Imitative Learning

Social motives in the training of manipulatory action are very import. Attending visually to the play-pal and the object the play-pal is demonstrating is crucial. Goals of the play-pal's actions and intentions must be considered. Sensitivity to such social stimuli as faces should be prioritized. When the iCub sees a face, it should activate attentional mechanisms for communication with and learning from the play-pal. There is an extensive literature on face perception in neonates and infants and it shows that visual sensitivity to faces and eye contact is innate. Furthermore, the ability to interpret gaze direction and pointing of the play pal must be considered.

Learning to Crawl

In addition to these scenarios, it is also intended that the iCub will learn to crawl. In this context, we will explore the possibility of sharing the same control circuitry for reaching with the forearms and for modulating the forearms during crawling (*e.g.* to do visually-guided hand placements).

In summary, our framework for the development of the iCub is as follows.

- The iCub starts with an innate visual-motor map that enables it to get the robot hand into the visual field. Thus, the robot also needs to have an innate conception of space in motor coordinates. We will investigate the possibility of developing this map, as described in Section 15.6.1, or deploying one that is pre-programmed. When the hand is in the visual field, the iCub tries to maintain it there. The iCub should also be able to move its hand towards graspable objects in the visual field. In order to do all this, the robot should be equipped with motives to move the hand into the visual field and towards objects that can be grasped. These motives will be based on some reward function such as the long-term decrease in entropy of some function of the iCub's behaviour, a decrease which may not be monotonic.

- When the robot can move the arm to the vicinity of objects in space, the visual system should begin to dock the hand onto the objects of interest. Certain anticipatory skills need to be built in to do this: the relationship between hand-orientation and the opposition spaces of objects, anticipation of when the object is encountered and a preparedness to grasp the object in preparation of this encounter. To begin with the object is grasped with the whole hand and the grasp is visually guided. Already at this developmental stage, the iCub should train to catch moving objects.

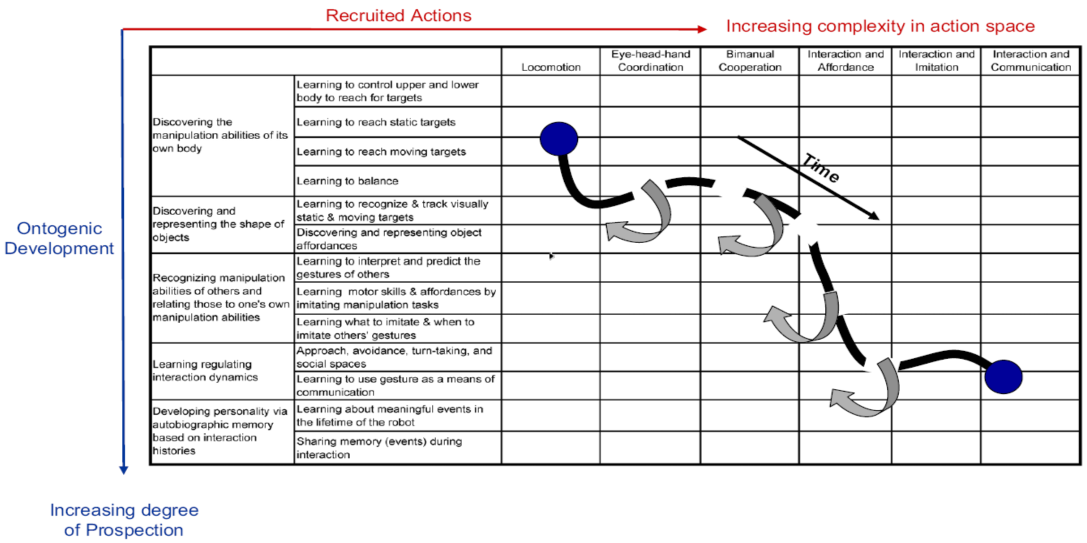
- The next step is to enable more exact control over the grasping action by controlling individual finger movements. In infants this occurs at around 9 months of age. The iCub will train to reach and grasp small artefacts like peas and objects of more complex forms. It will examine objects by squeezing, turning, and shaking them, and moving them from one hand to the other.

Once the iCub has mastered these skills, we will move on to experimental scenarios in which the iCub learns to develop object manipulation by playing on its own and or with another animate agent, that is, grasping objects and doing things in order to attain effects, like inserting objects into holes, building towers out of blocks *etc*. At his stage, social learning of object affordances becomes crucial. These scenarios will focus on the use of more than one object, emphasising the dynamic and static spatial relationships between them. In order of complexity, examples include:

- Learning to arrange block on a flat-surface;
- Learning to stack blocks of similar size and shape;
- Learning to stack blocks on similar shape but different size;
- Learning to stack blocks of different shape and size.

The chief point about these scenarios is they represent an opportunity for the iCub to develop a sense of spatial arrangment (both between itself and objects and between objects), and to arrange and order its local environment in some way. These scenarios also require that the iCub learn a set of primitive actions as well as their combination.

Figure 15 provides a diagrammatic summary of the overall strategy the project is adopting for the ontogenetic development of the iCub.


Recruited Actions
Increasing complexity in action space
Ontogenic Development
Increasing degree of Prospection
Time

		Locomotion	Eye-head-hand Coordination	Bimanual Cooperation	Interaction and Affordance	Interaction and Imitation	Interaction and Communication
Discovering the manipulation abilities of its own body	Learning to control upper and lower body to reach for targets						
	Learning to reach static targets						
	Learning to reach moving targets						
	Learning to balance						
Discovering and representing the shape of objects	Learning to recognize & track visually static & moving targets						
	Discovering and representing object affordances						
Recognizing manipulation abilities of others and relating those to one's own manipulation abilities	Learning to interpret and predict the gestures of others						
	Learning motor skills & affordances by imitating manipulation tasks						
	Learning what to imitate & when to imitate others' gestures						
Learning regulating interaction dynamics	Approach, avoidance, turn-taking, and social spaces						
	Learning to use gesture as a means of communication						
Developing personality via autobiographic memory based on interaction histories	Learning about meaningful events in the lifetime of the robot						
	Sharing memory (events) during interaction						

Figure 15: The iCub programme of ontogenetic development: increasingly prospective cognitive capabilities are developed over time by recruiting ever more complex actions.

17 Empirical Investigations

We have designed a series of experiments that investigate specific phylogenetic skills and ontogenetic development processes associated with the scenarios detailed above, especially the early ones. Since we wish to be as faithful as possible to natural development in humans, these investigations are a scripted version of the manner in which a psychologist would interact with a young infant during a series of typical sessions and they set out the behaviour that she or he would expect that infant to exhibit.

In these early experiments, we do not require the iCub to be able to re-position itself by crawling. Instead, the iCub sits in a special chair that gives support to the head and legs while the arms are free to move. We assume that the visual backdrop is a homogeneously coloured field and that the acoustic environment isn't noisy.

17.1 Looking

We begin by establishing the iCub's capabilites in *looking*.

1. *Saccades and gaze redirection*
   A face pattern is introduced into the peripheral visual field (30° from the centre). The visual angle corresponds to that of a real face at 0.5m. When this happens, the iCub moves the eyes and head to position the face at the centre of the visual field. They both start at the same time, but the eyes arrive first to its new position. When the eyes are at the final position and the head moves there, the gaze stays at the fixation object while the eyes counter rotate until they look straight ahead again. The same thing should also happen when a colourful object (3°– 8° visual angle) is introduced into the visual field or when a sounding object is introduced to the side of the robot (30° – 50°). New objects that the robot has not seen before will attract the gaze more than familiar objects.

2. *Gaze redirection and fixation*
   The robot turns its head (10° – 20°) while fixating an object or a face (10° – 30°). The eyes of the robot will then counter rotate so that the gaze is unaffected by the body movements (learning may be involved).

3. *Saccades, gaze redirection, and dynamic fixation (tracking)*
   An object moves into the visual field. Its average velocity is 8° – 25°/s. The robot makes a saccade to the object and then starts tracking it. The tracking will involve both head and eyes. When the object makes repetitive turns the robot should turn its eyes with the motion with no lag. When the turn is unexpected, a lag is acceptable but not greater than 0.1 seconds. The amplitude of the gaze adjustments may have smaller amplitude than the object motion and the difference will then be compensated with catch-up saccades to the object. Learning is involved. With training, the amplitude of the gaze adjustments will better adjusted to the object motion.

4. *Minimization of saccade correction by learning: tracking through occlusion*
   An object moves in the visual field and gets temporarily occluded behind some other objects. The robot stops the eyes at the disappearance point and then makes a saccade to the other side of the occluder. The saccade will predict when and where the object will appear.

A few notes are in order. First, it is clear that capability for smooth pursuit with prediction is required. Second, performance improvement by learning should be possible. Third, tracking through occlusion implies the modulation of (or action selection from) two distinct capabilities: smooth pursuit and saccade.

17.2 Reaching

We next proceed to address the iCub's ability in *reaching*. The situation is as above.

1. *Reaching towards a visual target (hand)*
   The robot extends one of its arms-hand into the visual field and then turns its head towards it. The robot will move the arm and try to keep its eyes on the hand all the time (again, learning may be involved in this). Both arms should be involved in this activity (first single limbs, then both limbs simultaneously). The robot should touch the other arm or hand when it is looking at it.

2. *Reaching towards a visual target (body)*
   The robot moves the arms to different parts of its own body and touches it. The hand opens up before or during the extension of the arm. This activity is carried out both when the robot looks at the different body parts and when it does not. The purpose of this activity is to build a body map (again, learning may be involved). The iCub will also touch body parts that lie outside the visual field.

3. *Reaching towards a visual target (moving object)*
   A ball or a cube (4-5 cm in diameter) is presented on a string or stick and gently moved up and down in front of the eyes. The robot turns the eyes and head towards it. It also extends one (or both) arms towards the object. The hand opens up during the extension of the arm and the fingers of the hand extends to make the touch surface larger. When the robot learns to reach, it might be an advantage to make the iCub always start the approach at a similar position. We have observed that the infants tend to retreat the hand closer to the body between attempts to get to the object but they do not seem to have a favourite lateral or vertical starting position. Another simplification of the reaching task is to lock the elbow joint. This has been reported in the literature but we have not observed it. It is possible that in special situations where the object is at a position where it can be attained without adjusting the elbow joint, the infant will only adjust the shoulder joint. When the hand of the robot touches the object, this activity will be repeated again and again with variation (that is, the robot retreats the hand a bit and makes a new approach) (again, learning). If the object is to the right, the right hand will be involved and if the object is to the left, the left hand will be involved. If the object is positioned straight ahead, one or both arms will extend towards it. Note that the focus of pre-reaching activity is on the arm. The hand acts as a feeler.

4. *Learning efficient reaching & learning when not to reach*
   The distance and lateral position to the ball or cube is varied from half the length of the arms to 1.5 the length of the arms. The iCub will learn to plan an efficient trajectory to the object. To begin with only a part of the trajectory will be planned ahead. At the end of this part, a new segment will be planned, *etc.* In the end, a continuous movement to the goal will be performed. If the distance to the object is larger than the arms, the robot will not reach for the object.

Again, some notes are in order. Turning the head toward the arm-hand as it enters the field of view is based on both visual and proprioceptive data. It implies a capability for hand detection and hand localization. The bimanual behaviour should be emergent. Moving the arm to different parts of the iCub body and touching them implies both haptic and force feedback. Note that the iCub is not yet equipped with haptic sensing.

17.3 Reach and Grasp

We now proceed to consider *reach and grasp*. The iCub sits independently.

1. *Reaching to a fixated static object*
   Objects of different sizes are introduced into the visual field of the iCub. The iCub extends one or both hands towards the object and then grasps it. The duration of the approach will be 3 seconds or less. The robot hand will slow down towards the end of the approach and just before grasping the object, the velocity will be close to zero. The iCub will fixate the object to be grasped during the approach.

2. *Grasp closure during approach*
   The hand will first open up during the approach of the object and then begin to close around it. All fingers will be engaged. To begin with, the hand will open to its full extent during the approach before starting to close. Later on during training, the maximal opening of the hand will be adjusted to the size of the object. The maximum opening of the hand should always be larger than the object to be grasped to make it easier to slide the hand over the object. It is important that the grasping begins before the touch otherwise there is a risk that the hand of the robot will push away the object as a consequence of the touch. The last part of the closing of the hand will take place as the iCub's hand is in contact with the object. If the object is large (< 10 cm diameter) both hands will participate in grasping the object. In order not to have the two hands compete for grasping the object, it might be desirable to develop some laterality.

3. *Matching grasp pose to an object's axis of symmetry*
   Objects of different forms are introduced into the visual field of the iCub (cylinders with a 2 cm and 5 cm diameter, and egg-shaped object with maximum diameter of 6 cm, an irregular object, and a soft and a hard object). The robot-hand will rotate during the approach in order to grasp the object over the most convenient opposition space. If the object is a rod, the grasp will take place around its longitudinal axis.

4. *Reaching to a fixated moving object*
   The object to be grasped moves. The velocity of the object motion will vary from 5 to 60 cm/s. The object will either approach on a vertical trajectory or a horizontal one. The hand moves towards a future position of the object where the hand and the object will meet. If the object comes from the left, it is the right arm-hand that will grasp it and if it comes from the right, it is the left arm-hand that will grasp it. The other hand will help to secure the object after the active hand has caught it (or stopped it).

5. *Pincer grasp*
   Small round objects (0.5 to 2.0 cm diameter) will be introduced into the visual field. The iCub will then only engage the thumb and the index finger in the act of grasping them.

6. *Bimanual manipulation and experimentation*
   After the object is grasped, the robot will examine the object by turning it around. Both hands will participate in this activity. One hand will hold the object in a fixed position while the other hand is moved over it in order to feel its surface and examine its interior.
   The iCub grasps the object and drops it on the floor while looking. The iCub picks it up again, rubs it on the floor and bangs it against the floor, tries to roll it, squeezes it, and moves it between the hands while looking. Through this activity the robot will build an object representation of familiar objects.

7. *Hand-to-hand transfer*
   The object will be transferred from one hand to the other while the robot fixates the object

(maybe also transferred repeatedly between the hands). The transfer should be as smooth and continuous as possible. This means that the delivering hand should let go of the object at the same time as the receiving hand grasps it.

8. *Hand and arm object relocation to a fixation point via intermediate landmarks*
   After grasping an object, the robot will move it to another position and deposit it there. The robot will turn its gaze towards the goal position of the action while the object is moved there. If the object is moved to its final position via an obstacle, the robot will fixate the obstacle and when the hand with the object has cleared the obstacle, the gaze will go to the final position.

Right hand reaching for objects on the right (and, similarly, left for those on the left) should not be pre-programmed but should be determined through action selection. The counterpart of this is that the right hand should reach for objects moving from the left (and vice versa, left reaching for those moving from the right). All of these behaviours should be a consequence of some predictive or anticipatory capability which modulates the action selection.

17.4 Reach and Posture

Once these capabilities have been demonstrated, we move on to consider *reaching and posture.* In this case, the iCub sits without support:

1. Exhibiting compensation for inertia and gravity,

2. Leaning forward,

3. and using the other hand to counterbalance.

17.5 Postural Control in Action

Similarly, the next stage in the development of the iCub deals with *postural control in action.* Here, the iCub sits independently and moves by crawling:

1. Crawls and prepares a reach during crawling. The iCub manages a transition from crawling to sitting.

2. Sitting and balancing.

3. Balancing during action. The iCub adjusts its posture: the body is stabilized so when the iCub grasps the other hand counter-balances.

17.6 Object Containment

The next stage is to consider *object containment*.

The iCub sits independently in front of two objects, one of them is smaller than the other which is larger and hollow. The smaller object can be fitted into the larger object.

1. The iCub picks up one of the objects and inspects it visually from several viewpoints. The iCub picks up the other object with the other hand and inspects it from several viewpoints. It then turns one of the object such that it fits into the other one.

17.7 Pointing and Gesturing

Finally, we consider *pointing and gesturing*.

The iCub sits in front of a human partner. An object is situated between them.

1. The iCub turns head and eyes toward the partner´s face and then towards the object and then towards the partner again. The iCub then opens the hand with the palm up and moves the upper body forward as if wanting the partner to give it the object.

17.8 A Comprehensive Experiment

The following experiment is designed to demonstrate the integration of all work-packages.

The robot is sitting in front of a human partner and there are two objects between them. The distance to the partner is 2 metres.

1. The iCub turns to look at one of the objects with head-eyes. It raises its right arm-hand and points to the attended object. It then assumes a crawling posture and crawls up to the objects. During the last stride the right arm is lifted (predictively).

2. When it arrives at the object, it assumes a sitting position, grasps the object and hands it to the human partner. This is repeated with the other object.

3. The human partner then picks up one of the objects and stretches it towards the iCub who opens the hand and grasps the object.

4. After this, the human partner picks up the other object and hands it to the iCub who transfers the first object to the other hand before receiving it.

5. Then the human partner turns his/her head and eyes toward one of the objects and points at it. The iCub turns its head and eyes toward the same object. The human partner then extends one of its arms, points to the object and places the hand in a begging posture. The iCub picks up the object and hands it to the human partner.

6. Now the human partner and the iCub have one object each. The human partner picks up his/her object and drops it into one of two buckets. After this the iCub picks up the other object and drops it into the other bucket (the gaze should move to the goal, *not* track the action).

References

[AAEM90] R. A. Andersen, C. Asanuma, G. Essick, and Siegel R. M. Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. J Comp Neurol., 296:65–113, 1990.

[AB02] A. Aguiar and R. Baillargeon. Developments in young infants’ reasoning about occluded objects. Cognitive Psychology, 45:267–336, 2002.

[ABB+04] 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.

[ADS86] G. E. Alexander, M. R. DeLong, and P. L. Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev. Neurosci., 9:357–381, 1986.

[AG89] R. A. Andersen and J. W. Gnadt. Role of posterior parietal cortex. In R. H. Wurtz and M. E. Goldberg, editors, The Neurobiology of Saccadic Eye Movements, Reviews of Oculomotor Research, volume 3, pages 315–335, Amsterdam, 1989. Elsevier.

[Ale90] I. Aleksander. Neural systems engineering: towards a unified design discipline? Computing and Control Engineering Journal, 1(6):259–265, 1990.

[And96] J. R. Anderson. Act: A simple theory of complex cognition. American Psychologist, 51:355–365, 1996.

[Ano64] P. K. Anokhin. Systemogenesis as a general regulator of brain development. Progress in Brain Research, 9, 1964.

[Ark98] A. Arkin. Behavior-based Robotics. MIT Press, Cambridge, MA, 1998.

[AS90] R. N. Aslin and S. L. Shea. Velocity thresholds in human infants: Implications for the perception of motion. Developmental Psychology, 26:589–598, 1990.

[Asl81] R. N. Aslin. Development of smooth pursuit in human infants. In D. F. Fisher, R. A. Monty, and J.W. Senders, editors, Eye Movements: Cognition and Visual Perception, Hillsdale N.J., 1981. Erlbaum.

[Atk00] J. Atkinson. The developing visual brain. Oxford University press, Oxford, UK, 2000.

[BA08] D. Badaley and K.E. Adolph. Beyond the average: Walking infants take steps longer than their leg length. Infant Behavior & Development, 31,554-558, 2008.

[Baa98] B. J. Baars. A Cognitive Theory of Consciousness. Cambridge University Press, 1998.

[Baa02] B. J. Baars. The conscious assess hypothesis: origins and recent evidence. Trends in Cognitive Science, 6(1):47–52, 2002.

[Bar93] G. R. Barnes. Visual-vestibular interaction in the control of head and eye movement: The role of visual feedback and predictive mechanisms. Progress in Neurobiology, 41(435-472), 1993.

[BASG95] P. R. Brotchie, R. A. Andersen, L. H. Snyder, and S. J. Goodman. Head position signals used by parietal neurons to encode locations of visual stimuli. Nature, 375: 232–235, 1995.

[Bat75] E. Bates, L. Camaioni, and V. Volterra. The adquisition of preformatives prior to speech. Merril-Palmer-Quarterly 21(3), 205-226, 1975.

[Bay69] N. Bayley. Bayley scales of infant development. Psychological Corporation, New York, 1969.

[BB78] A. J. Benson and G. R. Barnes. Vision during angular oscillations: The dynamic interaction of visual and vestibular mechanisms. Aviation Space and Environmental Medicine, 49:340–345, 1978.

[BB88] M. S. Banks and P. J. Bennett. Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. Journal of the Optical Society of America, 5:2059–2079, 1988.

[BBM+99] R. A. Brooks, C. Breazeal, M. Marajanovic, B. Scassellati, and M. M. Williamson. The cog project: Building a humanoid robot. In C. L. Nehaniv, editor, Computation for Metaphors, Analogy and Agends, volume 1562 of Springer Lecture Notes in Artificial Intelligence, Berlin, 1999. Springer-Verlag.

[BC92] H. Bloch and I. Carchon. On the onset of eye-head co-ordination in infants. Behavioural Brain Research, 49:85–90, 1992.

[BC95] S. Baron-Cohen. Mindblindness. MIT Press, Cambridge, MA, 1995.

[BCM82] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. Journal of Neurscience, 2(1):32–48, 1982.

[Ber67] N. Bernstein. The coordination and regulation of movements. Pergamon, Oxford, 1967.

[Ber96] N. E. Berthier. Learning to reach: a mathematical model. Developmental Psychology, 32:811–823, 1996.

[Ber00] A. Berthoz. The Brain's Sense of Movement. Harvard University Press, Cambridge, MA, 2000.

[BCG+96] N.E. Berthier, R. K. Clifton, V. Gullapalli, D. D. McCall, and D. J. Robin. Visual information and object size in the control of reaching. J. Mot Behav. 28(3), 187-197, 1996.

[B*etal*00] Barela et al., 2000

[BAD+] M. Barbu-Roth, D. Anderson, A. Despres, J. Provasi, and J. Campos. Neonatal Stepping to Terrestial Optic Flow. Child Devel., in press.

[BG06] B. Bertenthal and G. Gredeb¨ack. Information signifying occlusion in infants. In preparation, 06.

[BG85] C. J. Bruce and M. E. Goldberg. Primate frontal eye fields. I. single neurons discharging before saccades. J Neurophysiol., 53:603–635, 1985.

[BG87] R. Baillargeon and M. Graber. Where’s the rabbit? 5.5-month-old infants’ representation of the height of a hidden object. Cognitive Development, 2:375–392, 1987.

[BGH82] E. E. Birch, J. Gwiazda, and R. Held. Stereoacuity development for crossed and uncrossed disparities in human infants. Vision Research, 22:507–513, 1982.

[Bic00] M. H. Bickhard. Autonomy, function, and representation. Artificial Intelligence, Special Issue on Communication and Cognition, 17(3-4):111–131, 2000.

[Bil02] A. Billard. Imitation. In M. A. Arbib, editor, The Handbook of Brain Theory and Neural Networks, pages 566–569. MIT Press, Cambridge, MA, 2002.

[BJC99] Barela, Jeka, & Clark, 1999.

[BKS03] H. Barth, N. Kanwisher, and E. Spelke. The construction of large number representations in adults. Cognition, 86(201–221), 2003.

[BLL04] 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.

[BMS+05] 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.

[BN08] T. Barrett and A. Needham. Developmental differences in infants’ use of an object’s shape to grasp it securely. Developmental Psychobiology, 50, 97-106, 2008.

[BRB97] B. Bertenthal, J. Rose, and D. Bai. Perception-action coupling in the development of visual control of posture. Journal of Experimental Psychology: Human Perception and Performance, 23:1631–1643, 1997.

[Bre00] C. Breazeal. Sociable Machines: Expressive Social Exchange Between Humans and Robots. Unpublished Doctoral Dissertation. MIT, Cambridge, MA, 2000.

[Bre03] C. Breazeal. Emotion and sociable humanoid robots. International Journal of Human-Computer Studies, 59:119–155, 2003.

[Bro86] R. A. Brooks. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, RA-2(1):14–23, 1986.

[Bro02] R. A. Brooks. Flesh and Machines: How Robots Will Change Us. Pantheon Books, New York, 2002.

[But03] Butterworth, 2003.

[Byr03] 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.

[CCL+91] L. Camaioni, M. C. Caselli, E. Longbardi, and V. Volterra. A parent report instrument for early language assessment. First Language,11, 345-360, 1991.

[CFRU99] L. Craighero, L. Fadiga, G. Rizzolatti, and C. A. Umilta. Movement for perception: a motor-visual attentional effect. Journal of Experimenal Psychology: Human Perception and Performance., 1999.

[CGR89] C. Cavada and P. S. Goldman-Rakic. Posterior parietal cortex in rhesus monkey: II. evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. J Comp Neurol., 287:422–445, 1989.

[CH00a] W. D. Christensen and C. A. Hooker. An interactivist-constructivist approach to intelligence: self-directed anticipative learning. Philosophical Psychology, 13(1):5–45, 2000.

[CH00b] 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.

[CKL+04] 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.

[Cla94] H. H. Clark. Managing problems in speaking. Speech Communication, 15:243–250, 1994.

[Cla01] A. Clark. Mindware – An Introduction to the Philosophy of Cognitive Science. Oxford University Press, New York, 2001.

[CLM95] R. Carr´e, B. Lindblom, and P. MacNeilage. Rôle de l'acoustique dans l'évolution du conduit vocal humain. C. R. Acad. Sci. Paris, Tome 320(Série IIb), 1995.

[CM98] V. Corkum and C. Moore. The origins of joint visual attention in infants. Developmental Psychology, 34:28–38, 1998.

[CMA93] Clifton, R. K., Muir, D. W., Ashmead, D. H., & Clarkson, M. G. (1993). Is visually guided reaching in early infancy a myth? *Child Development, 64*(4), 1099-1110.

[CMD+90] M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, and S. E. Petersen. Attentional modulation of neural processing of shape, color, and velocity in humans. Science, 248:1556–9, 1990.

[CMD+91] M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, and S. E. Petersen. Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci., 11:2383–2402, 1991.

[CNF04] L. Craighero, M. Nascimben, and L. Fadiga. Eye position affects orienting of visuospatial attention. Current Biology, 14:331–333, 2004.

[CNT98] M. Carpenter, K. Nagell, and M. Tomasello. Social cognition,joint attention, and communicative competence from 9 to 15 months of age. Monogr Soc Res Child Dev., 63(4), i-vi, 1-143, 1998.

[CR04] L. Craighero and G. Rizzolatti. Annual review of physiology. 2004.

[CT96] D. Corbetta and E. Thelen. The developmental origins of bimanual coordination: a dyamic perspective. J Exp Psychol Hum Percept Perform, 22(2), 502-522, 1996.

[CVDD85] G. Chevalier, S. Vacher, J. M. Deniau, and M. Desban. Disinhibition as a basic process in the expression of striatal function. i. the striatonigral influence on tectospinal/tecto-diencephalic neurons. Brain Res, 334:215–226, 1985.

[DC85] J. M. Deniau and G. Chevalier. Disinhibition as a basic process in the expression of striatal functions. ii. the striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. Brain Res, (334):227–233, 1985.

[dF80] A. J. deCasper and W. P. Fifer. On human bonding: Newborns prefer their mothers' voices. Science, 208:1174–1176, 1980.

[DF89] J. L. Dannemiller and R. L. Freedland. The detection of slow stimulus movement in 2- to 5-month-old infants. Journal of Experimental Child Psychology, 47:337–355, 1989.

[DFP00] G. O. De´ak, R. A. Flom, and A. D. Pick. Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. Developmental Psychology, 36:511–523, 2000.

[DHM97] B. D'Entremont, S. M. J. Hains, and D.W. Muir. A demonstration of gaze following in 3- to 6-month-olds. Infant Behavior and Development, 20:569–572, 1997.

[DJ64] G. O. Dayton and M. H. Jones. Analysis of characteristics of fixation reflex in infants by use of direct current electro-oculography. Neurology, 14:1152–1156, 1964.

[Doy99] K. Doya. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Networks, 12:961–974, 1999.

[DPJ+94] J. Decety, D. Perani, M. Jeannerod, V. Bettinardi, B. Tadary, B. Woods, and J. C. Mazziotta. Mapping motor representations with PET. Nature, 371:600–602, 1994.

[dSS97] C. de Sperati and D. Stucchi. Recognizing the motion of a graspable object is guided by handedness. Neuroreport, 8:2761–2765, 1997.

[dVVP82] J. I. P. de Vries, G. H. A. Visser, and H. F. R. Prechtl. The emergence of fetal behaviour. i. qualitative aspects. Early Human Development, 23:159–191, 1982.

[FAS+80] Fox, R., Aslin, R. N., Shea, S. L., & Dumais, S. T. (1980). Stereopsis in human infants. Science, 207, 323 – 324.

[FMS+02] J. Fan, B. D. McCandliss, T. Sommer, A. Raz, andn M. I. Posner (2002). Testing the Efficiency and Independence of Attentional Networks, Journal of Cognitive Neuroscience, 14:3, 340 – 347.

[FCSJ02] T. Farroni, G. Csibra, F. Simeon, and M. H. Johnson. Eye contact detection in humans from birth. PNAS, 99:9602–9605, 2002.

[FD82] R. Fox and C. McDaniel. The perception of biological motion by human infants. Science, 218, 486-487, 1982.

[FDS04] L. Feigenson, S. Dehaene, and E. S. Spelke. Core systems of number. Trends in Cognitive Sciences, 8:307–314, 2004.

[Fet al.02] L. Feigenson and et al. The representations underlying infants' choice of more: object-files versus analog magnitudes. Psychological Science, 13:150–156, 2002.

[FFPR95] 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.

[FJ03] J. R. Flanagan and R. S. Johansson. Action plans used in action observation. Nature, 424(769–771), 2003.

[FL05] Fagard, J. & Lockman, J.J. (2005) The effect of task constraints on infants' (bi)manual strategy for grasping and exploring objects. Infant Behavior & Development, 28, 305 – 315.

[FSvH08] Fagard, J., Spelke, E.S., von Hofsten, C., (2008) Reaching and grasping a moving object in 6-, 8-, and 10-month olds:laterality aspects. Submitted manuscript

[FN99] W. J. Freeman and R. N´u˜nez. Restoring to cognition the forgotten primacy of action, intention and emotion. Journal of Consciousness Studies, 6(11-12):ix–xix, 1999.

[Fod83] J. A. Fodor. Modularity of Mind: An Essay on Faculty Psychology. MIT Press, Cambridge, MA, 1983.

[Fod00] J. A. Fodor. The Mind Doesn't Work that Way. MIT Press, Cambridge, MA, 2000.

[For85] H. Forssberg. Ontogeny of human locomotor control. i. infant stepping, supported locomotion, and transition to independent locomotion. Experimental Brain Research, 57:480–493, 1985.

[Fra77] S. Fraiberg. Insights from the blind. Basic Books, New York, 1977.

[GAFR96] S. T. Grafton, M. A. Arbib, L. Fadiga, and G. Rizzolatti. Localization of grasp representations in humans by PET: 2. observation compared with imagination. Experimental Brain Research, 112:103–111, 1996.

[Gal90] C. R. Gallistel. The organization of learning. MIT Press, Cambridge, Mass., 1990.

[Gar93] H. Gardner. Multiple Intelligences: The Theory in Practice. Basic Books, New York, 1993.

[GDG98a] M. Gentilucci, E. Daprati, and M. Gangitano. Implicit visual analysis in handedness recognition. Consciousness & Cognition, 7:478–493, 1998.

[GDG98b] M. Gentilucci, E. Daprati, and M. Gangitano. Right-handers and left-handers have different representations of their own hand. Cognitive Brain Research, 6:185–192, 1998.

[GFFR96] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. Brain, 119:593–609, 1996.

[GFL+88] M. Gentilucci, L. Fogassi, G. Luppino, M. Matelli, R. Camarda, and G. Rizzolatti. Functional organization of inferior area 6 in the macaque monkey. I. somatotopy and the control of proximal movements. Exp. Brain Res., 71:475–90, 1988.

[GHG97] M. S. Graziano, X. T. Hu, and C. G. Gross. Coding the locations of objects in the dark. Science, 277(239-41), 1997.

[Gib66] J. J. Gibson. The senses considered as perceptual systems. Houghton Mifflin, New York, 1966.

[GM98] S. Gillner and H. A. Mallot. Navigation and acquisition of spatial knowledge in a virtual maze. Journal of Cognitive Neuroscience, 10:445–463, 1998.

[GMB03] S. Goldin-Meadow and C. Butcher. Pointing towards two-word speech in young children. In S. Kita (ed). Pointing: where language, culture, and cognition meet. (pp 85-187). Mahwah, NJ. Erlbaum Ass., 2003.

[GP00] E. J. Gibson and A. Pick. An Ecological Approach to Perceptual Learning and Development. Oxford University Press, 2000.

[Gra99] G. H. Granlund. The complexity of vision. Signal Processing, 74:101–126, 1999.

[GS89] M. E. Goldberg and M. A. Segraves. The visual and frontal cortices. in: The neurobiology of saccadic eye movements. In R. H. Wurtz and M. E. Goldberg, editors, Reviews of Oculomotor Research, volume 3, pages 283–313, Amsterdam, 1989. Elsevier.

[GSPR83] M. Gentilucci, C. Scandolara, I. N. Pigarev, and G. Rizzolatti. Visual responses in the postarcuate cortex (area 6) of the monkey that are independent of eye position. Exp. Brain Res., 50:464–8, 1983.

[GvHB02] G. Gredeb¨ack, C. von Hofsten, and P. Boudreau. Infants' tracking of continuous circular motion and circular motion interrupted by occlusion. Infant Behaviour and Development, 144:1–21, 2002.

[GY84] C. E. Granrud and A. Yonas. Infants' perception of pictorially specified interposition. Journal of Experimental Child Psychology, 37(500–511), 1984.

[GYK97] 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.

[HABF96] M. Hadders-Algra, E. Brogren, and H. Forssberg. Ontogeny of postural adjustments during sitting in infancy: variation, selection and modulation. Journal of Physiology, 493:273–288, 1996.

[Hau82] 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.

[HBG80] R. Held, E. Birch, and J. Gwiazda. Stereoacuity in human infants. PNAS, 77:5572–5574, 1980.

[HD97] P. R. Huttenlocher and A. S. Dabholkar. Regional differences in synaptogenesis in human cerebral cortex. J. Comp. Neurol., 387:167–178, 1997.

[Heb49] D. O. Hebb. The Organization of Behaviour. John Wiley & Sons, New York, 1949.

[HVU00] V. Haehl, V. Vardaxis, and B. Ulrich. Learning to cruise: Bernstein's theory applied to skill acquisition during infancy. Human Movement Science, 19, 685-715, 2000.

[Hor01] I. Horswill. Tagged behavior-based systems: Integrating cognition with embodied activity. IEEE Intelligent Systems, pages 30–38, 2001.

[Hor06] I. Horswill. Cerebus: A higher-order behavior-based system. AI Magazine, 2006. In Press.

[HP74] J. Hyvarinen and A. Poranen. Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. Brain, 97:673–692, 1974.

[HRGfA92] L. Hainline, P. Riddell, J. Grose-fifer, and I. Abramov. Development of accommodation and convergence in infancy. Behavioural Brain Research, 49:30–50, 1992.

[HS96] L. Hermer and E. S. Spelke. Modularity and development: the case of spatial reorientation. Cognition, 61(195–232), 1996.

[Hut90] P. R. Huttenlocher. Morphometric study of human cerebral cortex development. Neuropsychologia, 28:517–527, 1990.

[HW83] O. Hikosaka and R. H. Wurtz. Visual and occulomotor functions of monkey substantia nigra pars reticulata. ill. memory contingent visual and saccade responses. J. Neurophysiol, 49:1268–1284, 1983.

[Hyé83] D. Hyén. The broad frequency-band rotary test, volume Dissertation Number 152. Linkoping University Medical School, Linkoping, Sweden, 83.

[IDC+05] Y. P. Ivanenko, N. Dominici, G. Cappellini, and F. Lacquantini. Kinematics in newly walking toddlers does not depend upon postural stability. J Neurophysiology,94,754-763, 2005.

[IDL07] Y. P. Ivanenko, N. Dominici, and F. Lacquantini. Development of independent walking in toddlers. Exerc.Sport. Sci.Rev.,35 82),67-73, 2007.

[Jet al.01] R. S. Johansson and et al. Eye-hand coordination in object manipulation. Journal of Neuroscience, 21:6917–6932, 2001.

[JM91] M. H. Johnson and J. Morton. Biology and cognitive development: the case of face recognition. Blackwell, Oxford, 1991.

[Joh00] S. H. Johnson. Thinking ahead: the case for motor imagery in prospective judgements of prehension. Cognition, 74:33–70, 2000.

[Jus92] P. W. Juscyck. Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon, editors, Phonological Development: Models, Research, Implications, pages 17–64, Timonium, Md, 1992. York Press.

[JV94] M. Jones and D. Vernon. Using neural networks to learn hand-eye co-ordination. Neural Computing and Applications, 2(1):2–12, 1994.

[KA98] P. J. Kellman and M. E. Arterberry. The cradle of knowledge. MIT Press, Cambridge, Mass., 1998.

[KdM00] Kayed and Van der Meer. Timing strategies used in defensive blinking to optical collisions in 5- to 7-month-old infants. Infant Behaviour and Development, 23:253–270, 2000.

[KE05] J. L. Krichmar and G. M. Edelman. Brain-based devices for the study of nervous systems and the development of intelligent machines. Artificial Life, 11:63–77, 2005.

[KE06] J. L. Krichmar and G. M. Edelman. Principles underlying the construction of brain-based devices. In T. Kovacs and J. A. R. Marshall, editors, Proceedings of AISB ’06 - Adaptation in Artificial and Biological Systems, volume 2 of Symposium on Grand Challenge 5: Architecture of Brain and Mind, pages 37–42, Bristol, 2006. University of Bristol.

[Kel95] J. A. S. Kelso. Dynamic Patterns – The Self-Organization of Brain and Behaviour. MIT Press, 3rd edition, 1995.

[KG06] O. Kochukhova and G. Gredeb¨ack. There are many ways to solve an occlusion task: the rolw of inertia and recent experience. Submitted Manuscript, 06.

[KS83] Kellman, P.J. & Spelke, E.S. (1983) Perception of partly occluded objects in infancy. Cognitive Psychology, 15, 483-524.

[Kih87] J. F. Kihlstrom. The cognitive unconscious. Science, 237:1445–1452, September 1987.

[Kit03] S. Kita (ed.) Pointing: Where language, culture and cognition meet. Mahwah, NJ: Erlbaum, 2003.

[KM97] 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.

[KNGE05] J. L. Krichmar, D. A. Nitz, J. A. Gally, and G. M. Edelman. Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. Proceedings of the National Academy of Science, USA, 102:2111–2116, 2005.

[KR05] J. L. Krichmar and G. N. Reeke. The darwin brain-based automata: Synthetic neural models and real-world devices. In G. N. Reeke, R. R. Poznanski, K. A. Lindsay, J. R. Rosenberg, and O. Sporns, editors, Modelling in the neurosciences: from biological systems to neuromimetic robotics, pages 613–638, Boca Raton, 2005. Taylor and Francis.

[KS83] P. J. Kellman and E. S. Spelke. Perception of partly occluded objects in infancy. Cognitive Psychology, 15:438–524, 1983.

[KS92] A. Karmiloff-Smith. Beyond Modularity: A developmental perspective on cognitive science. MIT Press, Cambridge, MA, 1992.

[KS94] A. Karmiloff-Smith. Precis of beyond modularity: A developmental perspective on cognitive science. Behavioral and Brain Sciences, 17(4):693–745, 1994.

[KSN+05] J. L. Krichmar, A. K. Seth, D. A. Nitz, J. G. Fleisher, and G. M. Edelman. Spatial navigation and causal analysis in a brain-based device modelling cortical-hippocampal interactions. Neuroinformatics, 3:197–221, 2005.

[Kuh94] P. K. Kuhl. Learning and representation in speech and language. Current opinion in Neurobiology, 4:812–822, 1994.

[Kuy73] H. G. J. M. Kuypers. The anatomical organization of the descending pathways and their contribution to motor control especially in primates. In J. E. Desmedt, editor, New Developments in Electromyography and Clinical Neurophysiology, 3, pages 38–68, New York, 1973. S. Karger.

[KvHvdWC90] P. J. Kellman, C. von Hofsten, van der Walle, and Condry. Perception of motion and stability during observer motion by pre-stereoscopic infants. In ICIS. Montreal, 1990.

[KVKD79] J. P. Kremenitzer, H. G. Vaughan, D. Kurtzberg, and K. Dowling. Smooth-pursuit eye movements in the newborn infant. Child Development, 50:442–448, 1979.

[LAB84] J. J. Lockman, D. H. Ashmead, and E.W. Bushnell. The development of anticipatory hand orientation during infancy. Journal of Experimental Child Psychology, 37:176–186, 1984.

[Lan04] 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.

[Lan05] 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.

[Lan06] P. Langley. Cognitive architectures and general intelligent systems. AI Magazine, 2006. In Press.

[Les94] A. M. Leslie. Tomm, toby, and agency: Core architecture and domain specificity. In L. A. Hirschfeld and S. A. Gelman, editors, Mapping the Mind: Specificity in Cognition and Culture, pages 119–148. Cambridge University Press, Cambridge, MA, 1994.

[Let al.84] B. Landau and et al. Spatial knowledge in a young blind child. Cognition, 16:225–260, 1984.

[Lew01] R. L. Lewis. Cognitive theory, soar. In International Encyclopedia of the Social and Behavioural Sciences, Amsterdam, 2001. Pergamon (Elsevier Science).

[LH76] D. G. Lawrence and D. A. Hopkins. The development of the motor control in the rhesus monkey: Evidence concerning the role of corticomotoneuronal connections. Brain, 99:235–254, 1976.

[LCS+06] U. Lizowski, M. Carpenter, T. Striano, and M. Tomasello. Twelve and 18 month-olds point to provide information Journal of Cognition and Development 7(2), 2006.

[LLR98] 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.

[LNR87] J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: an architecture for general intelligence. Artificial Intelligence, 33(1–64), 1987.

[Loc00] Lockman, J.J. (2000). A perception-action perspective on tool use development. Child Development, 71, 137 – 144.

[LW96] A. Ledebt and S. Wiener-Vacher. Head coordination in the sagittal plane in toddlers during walking:preliminary results. Brain Research Bulletin, 5/6, 371-373, 1996.

[MAB97] C. Moore, M. Angelopoulos, and P. Bennett. The role of movement in the development of joint visual attention. Infant Behavior and Development, 20:83–92, 1997.

[Mar77] D. Marr. Artificial intelligence – A personal view. Artificial Intelligence, 9:37–48, 1977.

[Mas03] N. Matasaka. From index finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In Pointing: Where language, culture and cognition meet, edited by S. Kita, 69-84. Mahwah, NJ: Erlbaum, 2003.

[Mat70] H. Maturana. Biology of cognition. Research Report BCL 9.0, University of Illinois, Urbana, Illinois, 1970.

[Mat75] H. Maturana. The organization of the living: a theory of the living organization. Int. Journal of Man-Machine Studies, 7(3):313–332, 1975.

[Mau85] D. Maurer. Infants' perception of faceness. In T. N. Field and N. Fox, editors, Social Perception in Infants, pages 37–66, Hillsdale, N. J., 1985. Lawrence Erlbaum Associates.

[MCGR86] M. Matelli, R. Camarda, M. Glickstein, and G. Rizzolatti. Afferent and efferent projections of the inferior area 6 in the macaque monkey. J Comp Neurol., 251:281–98, 1986.

[MCA+01] McCarty, M. E., Clifton, R. K., Ashmead, D. H., Lee, P., & Goubet, N. (2001). How infants use vision for grasping objects. Child Development, 72(4), 973-987.

[MD93] P. F. MacNeilage and B. L. Davis. Motor explanations of babbling and early speech patterns. In Boysson-Bardies et al., editor, Developmental Neurocognition: Speech and Face Processing in the First Year of Life, pages 341–352, Amsterdam, 1993. Kluwer Academic Publishers.

[Men83] L. Menn. Development of articulatory, phonetic and phonological capabilities. In B. Butterworth, editor, Language Production and Control, volume 2, pages 3–50, London, 1983. Academic Press.

[MMC+05] J. S. Metcalfe, K. McDowell, T.-Y. Chang, L.-C. Chen, J. J. Jeka, and J. E. Clark. Development of somatosensory-motor integration: an event-related analysis of infant posture in the first year of independent walking. Developmental Psychobiology, 46,19-35, 2005.

[MF03] G. Metta and P. Fitzpatrick. Early integration of vision and manipulation. Adaptive Behavior, 11(2):109–128, 2003.

[MG95] A. D. Milner and M. A. Goodale. The Visual Brain in Action. Oxford University Press, 1995.

[Mic04] O. Michel. Webots: professional mobile robot simulation. International Journal of Advanced Robotics Systems, 1(1):39–42, 2004.

[Min86] M. Minsky. Society of Mind. Simon and Schuster, New York, 1986.

[MLG+75] V. B. Mountcastle, J. C. G. A. Lynch, A. Georgopoulos, H. Sakata, and C. Acuna. Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. Journal of Neurophysiology, 38:871–908, 1975.

[MM77] A. N. Meltzoff and M. K. Moore. Imitation of facial and manual gestures by human neonates. Science, 198:75–78, 1977.

[MMD+00a] M. Morales, P. Mundy, C. E. F. Delgado, M. Yale, D. Messinger, and R. Neal. Responding to joint attention across the 6- through 24-month age period and early language acquisition. Journal of Applied Developmental Psychology, 21(283–298), 2000.

[MMD+00b] M. Morales, P. Mundy, C. E. F. Delgado, M. Yale, R. Neal, and H. K. Schwartz. Gaze following, temperament, and language development in 6-month-olds: A replication and extension. Infant Behavior and Development, 23:231–236, 2000.

[MMR98] M. Morales, P. Mundy, and J. Rojas. Following the direction of gaze and language development in 6-month-olds. Infant Behavior and Development, 21:373–377, 1998.

[MNO95] J. L. McClelland, B. L. NcNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological Review, 102(3):419–457, 1995.

[MRD95] P. Morissette, M. Ricard, and T. Gouin D'carie. Joint visual attention and pointing in infancy: A longitudinal study of comprehension. British Journal of Developmental Psychology, 13:163–175, 1995.

[MSK99] G. Metta, G. Sandini, and J. Konczak. A developmental approach to visually-guided reaching in artificial systems. Neural Networks, 12(10):1413–1427, 1999.

[MSS+05] J. G. McHaffie, T. R. Stanford, B. E. Stein, V. Coizet, and P. Redgrave. Subcortical loops through the basal ganglia. Trends in Neurosciences, 28(8):401–407, 2005.

[MV80] 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.

[MV87] H. Maturana and F. Varela. The Tree of Knowledge – The Biological Roots of Human Understanding. New Science Library, Boston & London, 1987.

[MW88] M. Müller and R. Wehner. Path integration in desert ants, cataglyphis fortis. PNAS, 85:5287–5290, 1988.

[Nan88] J. Nanez. Perception of impending collision in 3- to 6-week-old infants. Infant Behaviour and Development, 11:447–463, 1988.

[New82] A. Newell. The knowledge level. Artificial Intelligence, 18(1):87–127, March 1982.

[New90] A. Newell. Unified Theories of Cognition. Harvard University Press, Cambridge MA, 1990.

[NGPR99] J. Nadel, C. Guerini, A. Peze, and C. Rivet. The evolving nature of imitation as a format for communication. In J. Nadel and G. Butterworth, editors, Imitation in Infancy, pages 209–234. Cambridge University Press, Cambridge, 1999.

[NS76] A. Newell and H. A. Simon. Computer science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery, 19:113–126, March 1976. Tenth Turing award lecture, ACM, 1975.

[NSM+89] Newell K. M., Scully D. M., McDonald P. V., Baillargeon, R. (1989)Task constraints and infant grip configurations. Developmental Psychobiology, 22, 817-832.

[ODS02] B. Ogden, K. Dautenhahn, and P. Stribling. Interactional structure applied to the identification and generation of visual interactive behaviour: Robots that (usually) follow the rules. In I.Wachsmuth and T. Sowa, editors, Gesture and Sign Languages in Human-Computer Interaction, volume LNAI 2298 of Lecture Notes LNAI, pages 254–268. Springer, 2002.

[ONP06] L. Olsson, C. L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. Connection Science, 18(2), 2006.

[Pav90] M. Pavel. Predictive control of eye movement. In E. Kowler, editor, Eye Movements and Their Role in Visual and Cognitive Processes, volume 4 of Reviews of Oculomotor Research, pages 71–114, Amsterdam, 1990. Elsevier.

[PD94] M. I. Posner and S. Dehaene. Attentional networks. Trends Neurosci., 17:75–9, 1994.

[Pet al.02] T. J. Prescott and et al. The robot basal ganglia: action selection by an embedded model of the basal ganglia. In L. Nicholson and R. Faull, editors, Basal Ganglia VII, pages 349–356. Plenum Press, 2002.

[PFD+95] L. M. Parsons, P. T. Fox, J. H. Downs, T. Glass, T. B. Hirsch, C. C. Martin, P. A. Jerabek, and J. L. Lancaster. Use of implicit motor imagery for visual shape discrimination as revealed by PET. Nature, 375:54–58, 1995.

[Pia53] J. Piaget. The origins of intelligence in the child. Routledge, New York, 1953.

[Pia54] J. Piaget. The construction of reality in the child. Basic Books, New York, 1954.

[Pia55] J. Piaget. The Construction of Reality in the Child. Routeledge and Kegan Paul, London, 1955.

[Pin84] S. Pinker. Visual cognition: An introduction. Cognition, 18:1–63, 1984.

[Pin97] S. Pinker. How the Mind Works. W. W. Norton and Company, New York, 1997.

[PBG03] D. J. Pivonelli, J. M. Bering, and S. Giambrone. Chimpanzees' "pointing": Another error of the argument by analogy? In Pointing: Where language, culture and cognition meet, edited by S. Kita, 33-68. Mahwah, NJ: Erlbaum, 2003.

[PP84] M. Petrides and D. N. Pandya. Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. J Comp Neurol., 228:105–16, 1984.

[PP90] M. I. Posner and S. E. Petersen. The attention system of the human brain. Annu Rev Neurosci., 13:25–42, 1990.

[PPFR88] M. I. Posner, S. E. Petersen, P. T. Fox, and M. E. Raichle. Localization of cognitive operations in the human brain. Science, 240(1627-31), 1988.

[Pyl84] Z. W. Pylyshyn. Computation and Cognition. Bradford Books, MIT Press, 2^{nd} edition, 1984.

[Ram85] D. S. Ramsay. Fluctuations in unimanual hand preference in infants following the onset of duplicated syllable babbling. Developmental Psychology, 21, 318-324, 1985.

[RBC96] Robin, D. J., Berthier, N. E., & Clifton, R. K. (1996). Infants' predictive reaching for moving objects in the dark. Developmental Psychology, 32, 824-835.

[RC87] G. Rizzolatti and R. Camarda. Neural circuits for spatial attention and unilateral neglect. In M. Jeannerod, editor, Neurophysiological and neuropsychological aspects of spatial neglect, pages 289–313, Amsterdam, 1987. North-Holland.

[Ree96] E. S. Reed. Encountering the world: towards an ecological psychology. Oxford University Press, New York, 1996.

[R*etal*95] L. Regolin and et al. Object and spatial representations in detour problems by chicks. Animal Behaviour, 49:195–199, 1995.

[RNR+08] L. Righetti, A. Nyle´n, K. Rosander, and A. Ijspeert. Is the locomotion of crawling infants different from other quadropeded mammals? Submitted, 2008.

[RFFG97] G. Rizzolatti, L. Fadiga, L. Fogassi, and V. Gallese. The space around us. Science, pages 190–191, 1997.

[RFG97] G. Rizzolatti, L. Fogassi, and V. Gallese. Parietal cortex: from sight to action. Current Opinion in Neurobiology, 7:562–567, 1997.

[RFGF96] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi. Premotor cortex and the recognition of motor actions. Cognitive Brain Research, 3:131–141, 1996.

[RG95] P. Rochat and N. Goubet. Development of sitting and reaching in 5- to 6-month-old infants. Infant Behavior and Development, 18:53–68, 1995.

[RGNvH06] K. Rosander, G. Gredeb¨ack, P. Nystro¨om, and C. von Hofsten. Submitted manuscript. 2006.

[RH96] S. Retaux and W. A. Harris. Engrailed and retinotectal topography. Trends in Neuroscience, 19:542–546, 1996.

[RLN93] P. Rosenbloom, J. Laird, and A. Newell, editors. The Soar Papers: Research on Integrated Intelligence. MIT Press, Cambridge, Massachusetts, 1993.

[Roc92] P. Rochat. Self-sitting and reaching in 5- to 8-month-old infants: the impact of posture and its development on early eye-hand coordination. Journal of Motor Behavior, 24:210–220, 1992.

[Rou01] N. P. Rougier. Hippocampal auto-associative memory. In International Joint Conference on Neural Networks, 2001.

[Roz76] P. Rozin. The evolution of intelligence and access to cognitive unconscious. Psychobiology and Physiological Psychology, 6:245–279, 1976.

[RRDC87] G. Rizzolatti, L. Riggio, I. Dascola, and Umilta C. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia, 25:31–40, 1987.

[RRS94] G. Rizzolatti, L. Riggio, and B. M. Sheliga. Space and selective attention. In C. Umilt`a and M. Moscovitch, editors, Attention and performance XV, pages 231–265, Cambridge, MA, 1994. MIT Press.

[RS99] Rochat, P. And Striano T. (1999) Socio-emotional development in the first year of life. In P. Rochat (Ed.), Early social cognition. Mahwah, N.J.: Erlbaum.

[RvH00] K. Rosander and C. von Hofsten. Visual-vestibular interaction in early infancy. Exp. Brain Res., 133:321–333, 2000.

[RvH04] K. Rosander and C. von Hofsten. Infants' emerging ability to represent object motion. Cognition, 91:1–22, 2004.

[RY01] 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.

[SB75] M. Scaife and J. S. Bruner. The capacity for joint visual attention in infants. Nature, 53:265–266, 1975.

[SB90] M. J. Swain and D. H. Ballard. "Indexing via colour histograms", pp. 390–393, 1990.

[SB91] M. Swain and D. Ballard. "Color indexing". Internation Journal of Computer Vision, 7(1):11–32, 1991.

[SB05] M. P. Shanahan and B. Baars. Applying global workspace theory to the frame problem. Cognition, 98(2):157–176, 2005.

[Sca02] B. Scassellati. Theory of mind for a humanoid robot. Autonomous Robots, 12:13–24, 2002.

[SDC+96] A. Sirigu, J. R. Duhamel, L. Cohen, B. Pillon, B. Dubois, and Y. Agid. The mental representation of hand movements after parietal cortex damage. Science, 273:1564–1568, 1996.

[SOH08] W. Sanefuji, H. Ohgami, and K. Hashiya. Detection of the relevant type of locomotion in infancy:crawlers versus walkers. Infant Behav and Development, in press.

[Sha92] C. J. Shatz. The developing brain. Scientific American, pages 35–41, September 1992.

[Sha05a] M. P. Shanahan. Cognition, action selection, and inner rehearsal. In Proceedings IJCAI Workshop on Modelling Natural Action Selection, pages 92–99, 2005.

[Sha05b] 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.

[Sha06] M. P. Shanahan. A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition, 2006. To Appear.

[Sid95] Siddiqui, A. (1995) Object size as a determinant of grasping in infancy. Journal of Genetic Psychology.156: 345-58.

[SJS02] Smith,W.C., Johnson,S.P., Spelke, E.S. (2002) Motion and edge sensitivity in perception of object unity. Cognitive Psychology, 46, 31-64.

[SME+04] A.K. Seth, J.L. McKinstry, G.M. Edelman, , and J. L Krichmar. Active sensing of visual and tactile stimuli by brain-based devices. International Journal of Robotics and Automation, 19(4):222–238, 2004.

[SMV04a] G. Sandini, G. Metta, and D. Vernon. Robotcub: An open framework for research in embodied cognition. In IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids 2004), pages 13–32, 2004.

[SMV04b] G. Sandini, G. Metta, and D. Vernon. Robotcub: An open research initiative in embodied cognition. In Third International Conference on Development and Learning (ICDL ‘04), 2004.

[Spe89] G. S. Speidel. Imitation: a bootstrap for learning to speak. In G. E. Speidel and K. E. Nelson, editors, The many faces of imitation in language learning, pages 151–180. Springer Verlag, 1989.

[Spe90] Spelke, E.S (1990) Principles of object perception. Cognitive Science, 14, 29-56.

[SvdW93] E. S. Spelke and van der Walle. Perceiving and reasoning about objects: In sights from infants. In N. Eilan, R. McCarthy, & W. Brewer (Eds.) Spatial Representation. Oxford: Blackwell, 1993.

[Spe00] E. S. Spelke. Core knowledge. American Psychologist, pages 1233–1243, November 2000.

[Spe03] E. S. Spelke. Core knowledge. In N. Kanwisher and J. Duncan, editors, Attentionand Performance, volume 20. Oxford University Press, 2003.

[SR99] T. Striano and P. Rochat. Socio-emotional development in the first year of life. In P. Rochat, editor, Early social cognition, Mahwah, N.J., 1999. Erlbaum.

[SRCR95a] B. M. Sheliga, L. Riggio, L. Craighero, and G. Rizzolatti. Spatial attention and eye movements. Exp Brain Res, 105:261–75, 1995.

[SRCR95b] B. M. Sheliga, L. Riggio, L. Craighero, and G. Rizzolatti. Spatial attention determined modifications in saccade trajectories. Neuroreport, 6:585–8, 1995.

[SSG90] P. Starkey, E. S. Spelke, and R. Gelman. Numerical abstraction by human infants. Cognition, 36:97–127, 1990.

[ST01] T. Striano and M. Tomasello. Infant development: Physical and social cognition. In International encyclopedia of the social and behavioral sciences, pages 7410–7414, 2001.

[SvHK89] E. S. Spelke, C. von Hofsten, and R. Kestenbaum. Object perception and object directed reaching in infancy: interaction of spatial and kinetic information for object boundaries. Developmental Psychology, 25:185–196, 1989.

[TCS96] Thelen, E., Corbetta, D., & Spencer, J. P. (1996). Development of reaching during the first year: Role of movement speed. Journal of Experimental Psychology: Human Perception and Performance, 22, 1059-1076.

[The95] 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.

[TKF99] C. Trevarthen, T. Kokkinaki, and G. A. Fiamenghi Jr. What infants' imitations communicate: with mothers, with fathers and with peers. In J. Nadel and G. Butterworth, editors, Imitation in Infancy, pages 61–124. Cambridge University Press, Cambridge, 1999.

[TLB92] S. P. Tipper, C. Lortie, and G. C. Baylis. Selective reaching: evidence for action centered attention. J Exp Psychol Hum Percept Perform, 18:891–905, 1992.

[TFR84] E. Thelen, D. M. Fischer, and R. Ridley-Johnson. The relationship between physical growth and a newborn reflex. Infant Behavior and Development, 7, 479-493, 1984.

[TS94] 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.

[TS03] E. Thelen and L. B. Smith. Development as a dynamic system. Trends Cognitive Science, 7:343–348, 2003.

[Tom06] M. Tomasello. Acquiring linguistic constructions. In D. Kuhn & R. Siegler (Eds.), Handbook of Child Psychology. New York: Wiley, 2006.

[TZ93] Thelen, E., Corbetta ,D., Kamm,K., Spencer, I.P., Schneider, K. and R. F. Zernicker. The transition to reaching: Mapping intention and intrinsic dynamics. Child Development, 64:1058–1099, 1993.

[Uet al.01] M. A. Umilta and et al. I know what you are doing: A neurophysiological study. Neuron, 31(155–165), 2001.

[UM82] L. G. Ungerleider and M. Mishkin. Two visual systems. In D. J. Ingle, M. A. Goodale, and Mansfield R. J. W., editors, Analysis of visual behavior, pages 549– 586, Cambridge, MA, 1982. MIT Press.

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

[Var79] F. Varela. Principles of Biological Autonomy. Elsevier North Holland, New York, 1979.

[Var92] 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.

[vdMet al.96] A. H. L. van der Meer and et al. Lifting weights in neonates: developing visual control of reaching. Scandinavian Journal of Psychology, 37:424–436, 1996.

[vdMS88] C. von der Malsburg and W. Singer. Principles of cortical network organisations. In P. Rakic and W. Singer, editors, Neurobiology of the Neocortex, pages 69–99, London, 1988. John Wiley & Sons Ltd.

[vdM+95] A. L. H. van der Meer, F. R. van derWeel, and D. N. Lee. The functional significance of arm movements in neonates. Science, 267:693–695, 1995.

[vE+08] M. Van Elk, H. T. van Schie, S. Hunnius, C. Vesper, and H. Bekkering. You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. Neuroimage, in press.

[Ver06] D. Vernon. The space of cognitive vision. In H. I. Christensen and H.-H. Nagel, editors, Cognitive Vision Systems: Sampling the Spectrum of Approaches, LNCS, pages 7–26, Heidelberg, 2006. Springer-Verlag.

[Ver07] D. Vernon. Cognitive vision: The case for embodied perception. Image and Vision Computing, In Press:1–14, 2007.

[vH79] C. von Hofsten. Development of visually guided reaching: the approach phase. Journal of Human Movement Studies, 5:160–178, 1979.

[vH80] C. von Hofsten. Predictive reaching for moving objects by human infants. Journal of Experimental Child Psychology, pages 369–382, 1980.

[vH82a] C. von Hofsten. Binocular convergence as a determinant of reaching behaviour in infancy. Perception, 6:139–144, 1982.

[vH82b] C. von Hofsten. Eye-hand coordination in newborns. Developmental Psychology, 18:450–461, 1982.

[vH83] C. von Hofsten. Catching skills in infancy. Experimental Psychology: Human Perception and Performance, 9:75–85, 1983.

[vH84] C. von Hofsten. Developmental changes in the organization of pre-reaching movements. Developmental Psychology, 20:378–388, 1984.

[vH86] C. von Hofsten. The early development of the manual system. In B. Lindblom and R. Zetterstr¨om, editors, Precursors of Early Speech, Basingstoke, Hampshire, 1986. Macmillan.

[vH91] C. von Hofsten. Structuring of early reaching movements: A longitudinal study. Journal of Motor Behavior, 23:280–292, 1991.

[vH93] C. von Hofsten. Prospective control: A basic aspect of action development. Human Development, 36:253–270, 1993.

[vH97] C. von Hofsten. On the early development of predictive abilities. In C. Dent and P. Zukow-Goldring (Eds.) Evolving Explanations of Development: Ecological approaches to Organism-Environmental Systems. pp. 163-194, 1997.

[vH04] C. von Hofsten. An action perspective on motor development. Trends in Cognitive Science, 8:266–272, 2004.

[vHDF05] C. von Hofsten, E. Dahlstrom, and Y. Fredriksson. 12-month–old infants' perception of attention direction in static video images. Infancy, in press, 2005.

[vHFS00] C. von Hofsten, Q. Feng, and E. S. Spelke. Object representation and predictive action in infancy. Developmental Science, 3:193–205, 2000.

[vHFZ84] C. von Hofsten and S. Fazel-Zandy. Development of visually guided hand orientation in reaching. Journal of Experimental Child Psychology, 38:208–219, 1984.

[vHJ05] C. von Hofsten and K. Johansson. Planning to reach for a rotating rod: Developmental aspects. Manuscript, 2005.

[vHKP92] C. von Hofsten, P. J. Kellman, and J. Putaansuu. Young infants' sensitivity to motion parallax. Infant Behaviour and Development, 15:245–264, 1992.

[vHKR06] C. von Hofsten, O. Kochukhova, and K. Rosander. Predictive occluder tracking in 4-month-old infants. Submitted Manuscript, 06.

[vHL79] C. von Hofsten and K. Lindhagen. Observations on the development of reaching for moving objects. Journal of Experimental Child Psychology, 28:158–173, 1979.

[vHO05] C. von Hofsten and H. Örnkloo. Fitting objects into holes: The development of spatial cognition skills. Submitted, 2005.

[vHO09] C. von Hofsten. Action, the foundation for cognitive development. Scandinavian Journal of Psychology, 51, 1-7, 2009.

[vHR88] C. von Hofsten and L. Rönnqvist. Preparation for grasping an object: A developmental study. Journal of Experimental Psychology: Human Perception and Performance, pages 610–621, 1988.

[vHR96] C. von Hofsten and K. Rosander. The development of gaze control and predictive tracking in young infants. Vision Research, 36:81–96, 1996.

[vHR97] C. von Hofsten and K. Rosander. Development of smooth pursuit tracking in young infants. Vision Research, 37:1799–1810, 1997.

[vHS85] C. von Hofsten and E. S. Spelke. Object perception and object directed reaching in infancy. Journal of Experimental Psychology: General, 114:198–212, 1985.

[vHVS+98] von Hofsten, C., Vishton, P., Spelke, E.S., Feng, Q., and Rosander, K. (1998) Predictive action in infancy: Tracking and reaching for moving objects. Cognition, 67, 255-285.

[vHW90] C. von Hofsten and M. Woollacott. Postural preparations for reaching in 9-month old infants. Unpublished data, 1990.

[VMS09] D. Vernon, G. Metta, and G. Sandini. Embodiment in Cognitive Systems: on the Mutual Dependence of Cognition & Robotics, Invited Chapter to appear in "Embodied Cognitive Systems", J. Gray and S. Nefti-Meziani (Eds.), Institution of Engineering and Technology (IET), UK, 2009.

[VSM07] D. Vernon, G. Sandini, and G. Metta. The icub cognitive architecture: Interactive development in a humanoid robot. In Proceedings of IEEE International Conference on Development and Learning (ICDL), Imperial College, London, 2007.

[Vyg78] L. Vygotsky. Mind in society: The development of higher psychological processes. Cambridge, M.A.: Harvard University Press, 1978.

[WDMM87] M. Woollacott, M. Debu, M., and Mowatt. Neuromuscular control of posture in the infant and child: Is vision dominant? Journal of Motor Behavior, 19:167–186, 1987.

[Wen02] 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.

[Wen04a] J.Weng. Developmental robotics: Theory and experiments. International Journal of Humanoid Robotics, 1(2):199–236, 2004.

[Wen04b] J. Weng. A theory of developmental architecture. In Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004), La Jolla, October 2004.

[W*etal*02] D. C. Witherington and et al. The development of anticipatory postural adjustments in infancy. Infancy, 3:495–517, 2002.

[WF86] T. Winograd and F. Flores. Understanding Computers and Cognition – A New Foundation for Design. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1986.

[WG02] A. L.Woodward and J. J. Guajardo. Infants' understanding of the point gesture as an object-directed action. Cognitive Development, 17:1061–1084, 2002.

[WHZ+00] J. Weng, W. Hwang, Y. Zhang, C. Yang, and R. Smith. Developmental humanoids: Humanoids that develop skills automatically. In Proceedings the first IEEE-RAS International Conference on Humanoid Robots, Cambridge, MA, 2000.

[Wol87] P. H. Wolff. The development of behavioral states and the expression of emotions in early infancy. Chicago University Press, 1987.

[WWC+07] T. Wilcox, R. Woods, C. Chapa, and S. McCurry. Multisensory exploration and object individuation in infancy. Developmental Psychology, 43, 479-495, 2007.

[Wit08] D. C. Witherington. The development of prospective grasping control between 5 and 7 months: A longitudinal study. Infancy (in press).

[WS02] R. F. Wang and E. S. Spelke. Human spatial representation: insights from animals. Trends in Cognitive Sciences, 6:376–382, 2002.

[WTF86] A. M. Wing, A. Turton, and C. Fraser. Grasp size and accuracy of approach in reaching. Journal of motor Behavior, 18:245–261, 1986.

[Wyn92] K. Wynn. Addition and subtraction in infants. Nature, 358:749–750, 1992.

[WZ02] J. Weng and Y. Zhang. Developmental robots - a new paradigm. In Proc. Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, 2002.

[XS00] F. Xu and E. S. Spelke. Large number discrimination in 6-month-old infants. Cognition, 74:B1–B11, 2000.

[YAG87] A. Yonas, M. E. Arterberry, and C. E. Granrud. Space perception in infancy. In Vasta, editor, Annals of Child Development, pages 1–34, Greenwich, CT, 1987. JAI Press.

[YPL79] A. Yonas, L. Pettersen, and J. J. Lockman. Infants' sensitivity to optical information for collision. Canadian Journal of Psychology, 33:268–276, 1979.

[Zie01] T. Ziemke. Are robots embodied? In Balkenius, Zlatev, Dautenhahn, Kozima, and Breazeal, editors, Proceedings of the First International Workshop on Epigenetic Robotics—Modeling Cognitive Development in Robotic Systems, volume 85 of Lund University Cognitive Studies, pages 75–83, Lund, Sweden, 2001.

[Zie03] T. Ziemke. What's that thing called embodiment? In Alterman and Kirsh, editors, Proceedings of the 25th Annual Conference of the Cognitive Science Society, Lund University Cognitive Studies, pages 1134–1139, Mahwah, NJ, 2003. Lawrence Erlbaum.

[ZBD+07] S. Zoia, L. Blasen, G. DÒttavio, M. Bulgheroni, E. Pezzetta, A. Scatar, and U. Castielo. Evidence of early of action planning in the human foetus: a kinematic study. EBR, 176, 217-226, 2007.