11

Vernon et al.

A Roadmap for Cognitive Development in Humanoid Robots

David Vernon
Claes von Hofsten
Luciano Fadiga

»COSMOS 11

»COGNITIVE SYSTEMS MONOGRAPHS

A Roadmap for Cognitive Development in Humanoid Robots

Springer

Cognitive Systems Monographs

Volume 11

Editors: Rüdiger Dillmann · Yoshihiko Nakamura · Stefan Schaal · David Vernon

David Vernon, Claes von Hofsten,
and Luciano Fadiga

A Roadmap for Cognitive Development in Humanoid Robots

Springer

Rüdiger Dillmann, University of Karlsruhe, Faculty of Informatics, Institute of Anthropomatics, Humanoids and Intelligence Systems Laboratories, Kaiserstr. 12, 76131 Karlsruhe, Germany

Yoshihiko Nakamura, Tokyo University Fac. Engineering, Dept. Mechano-Informatics, 7-3-1 Hongo, Bukyo-ku Tokyo, 113-8656, Japan

Stefan Schaal, University of Southern California, Department Computer Science, Computational Learning & Motor Control Lab., Los Angeles, CA 90089-2905, USA

David Vernon, Department of Robotics, Brain and Cognitive Sciences, Italian Institute of Technology, Genoa, Italy

Authors

David Vernon
Department of Robotics, Brain and Cognitive Sciences
Italian Institute of Technology
Genoa
Italy
E-mail: david@vernon.eu

Claes von Hofsten
Psykologisk institutt
Universitetet i Oslo
Oslo
Norway
E-mail: claes.von_hofsten@psyk.uu.se

Luciano Fadiga
Section of Human Physiology
University of Ferrara
Italy
E-mail: fdl@unife.it

and

Department of Robotics, Brain and Cognitive Sciences
Italian Institute of Technology
Genoa

ISBN 978-3-642-16903-8 e-ISBN 978-3-642-16904-5

DOI 10.1007/978-3-642-16904-5

Cognitive Systems Monographs ISSN 1867-4925

Library of Congress Control Number: 2010938643

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

5 4 3 2 1 0

springer.com

Preface

The work described in this book is founded on the premise that (a) cognition is the process by which an autonomous self-governing agent acts effectively in the world in which it is embedded, that (b) the dual purpose of cognition is to increase the agent's repertoire of effective actions and its power to anticipate the need for and outcome of future actions, and that (c) development plays an essential role in the realization of these cognitive capabilities.

Cognitive agents act in their world, typically with incomplete, uncertain, and time-varying sensory data. The chief purpose of cognition is to enable the selection of actions that are appropriate to the circumstances. However, the latencies inherent in the neural processing of sense data are often too great to allow effective action. Consequently, a cognitive agent must anticipate future events so that it can prepare the actions it may need to take. Furthermore, the world in which the agent is embedded is unconstrained so that it is not possible to predict all the circumstances it will experience and, hence, it is not possible to encapsulate *a priori* all the knowledge required to deal successfully with them. A cognitive agent then must not only be able to anticipate but it must also be able to learn and adapt, progressively increasing its space of possible actions as well as the time horizon of its prospective capabilities. In other words, a cognitive agent must develop.

There are many implications of this stance. First, there must be some starting point for development — some phylogeny — both in terms of the initial capabilities and in terms of the mechanisms which support the developmental process. Second, there must be a developmental path — an ontogeny — which the agent follows in its attempts to develop an increased degree of prospection and a larger space of action. This involves several stages, from coordination of eye-gaze, head attitude, and hand placement when reaching, through to more complex 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. Third, since cognitive agents rarely operate in isolation and since the world with which they interact typically includes other cognitive agents, there is the question of how a cognitive agent can share with other agents the knowledge it has learned. Since what an agent knows is based on its history of experiences in the world, the meaning of any shared knowledge depends on a common mode of experiencing the world. In turn, this implies that the shared knowledge is predicated upon the agents having a common morphology and, in the case of human-robot interaction, a common humanoid form.

The roadmap set out in this book targets specifically the development of cognitive capabilities in humanoid robots. It identifies the necessary and hopefully sufficient conditions that must exist to allow this development. It has been created by bringing together insights from four areas: enactive cognitive science, developmental psychology, neurophysiology, and cognitive modelling. Thus, the roadmap builds upon what is known about the development of natural cognitive systems and what is known about computational modelling of artificial cognitive agents. In doing so, it identifies the essential principles of a system that can develop cognitive capabilities and it shows how these principles have been applied to the state-of-the-art humanoid robot: the iCub.

The book is organized as follows. Chapter 1 presents a conceptual framework that forms the foundation of the book. It identifies the broad stance taken on cognitive systems — emergent embodied systems that develop cognitive skills as a result of their action in the world — and it draws out explicitly the consequences of adopting this stance. Chapter 2 begins by discussing the importance of action as the organizing principle in cognitive behaviour, a theme that will recur repeatedly throughout the book. It then addresses the phylogeny of human infants and, in particular, it considers the innate capabilities of pre-natal infants and how these develop before and just after birth. Chapter 3 then details how these capabilities develop in the first couple of years of life, focussing on the interdependence of perception and action. In doing so, it develops the second recurrent theme of the book: the central role of prospective capabilities in cognition. Chapter 4 explores the neurophysiology of perception and action, delving more deeply into the way that the interdependency of perception and action is manifested in the primate brain. While Chapters 2 – 4 provide the biological inspiration for the design of an entity that can develop cognitive capabilities, Chapter 5 surveys recent attempts at building artificial cognitive systems, focussing on cognitive architectures as the basis for development. Chapter 6 then presents a complete roadmap that uses the phylogeny and ontogeny of natural systems as well as insights gained from computational models and cognitive architectures to define the innate capabilities with which the humanoid robot must be equipped so that it is capable of ontogenetic development. The roadmap includes a series of scenarios that can be used to drive the robot's developmental progress. Chapter 7 provides an overview of the iCub humanoid robot and it describes the use of the the roadmap in the realization of the iCub's own cognitive architecture. Chapter 8 concludes by setting out an agenda for future research and addressing the most pressing issues that will advance our understanding of cognitive systems, artificial and natural.

Dublin, Uppsala, and Ferrara
August 2010

David Vernon
Claes von Hofsten
Luciano Fadiga

Acknowledgements

This book is based on the results of the RobotCub research project, the goal of which was to develop the iCub: an open and widely-adopted humanoid robot for cognitive systems research. This project was supported by the European Commission, Project IST-004370, under Strategic Objective 2.3.2.4: Cognitive Systems and we gratefully acknowledge the funding that made this research possible.

In particular, we would like to thank Hans-Georg Stork, European Commission, for his support, insight, and unfailing belief in the project.

We would also like to acknowledge the contributions made by the five reviewers of the project over its five-and-a-half year lifetime: Andreas Birk, Joanna Bryson, Bob Damper, Peter Ford Dominey, and Raul Rojas. Their collective suggestions helped greatly in navigating uncharted waters.

We are indebted to the members of the project's International Advisory Board — Rodney Brooks, MIT, Gordon Cheng, Technical University of Munich, Jürgen Konczak, University of Minnesota, Hideki Kozima, CRL, and Yasuo Kuniyoshi, University of Tokyo — for providing valuable advice and moral support.

Over one hundred people were involved in the creation of the iCub and it is impossible to acknowledge the contribution each made to the work that is described in this book. However, certain key individuals in each of the eleven institutes that participated in the research played a pivotal role is bringing the five year project to a successful conclusion. It is a pleasure to acknowledge their contributions.

Giulio Sandini, Italian Institute of Technology and University of Genoa, was the mastermind behind the project and he was the first to see the need for a common humanoid robot platform to support research in embodied cognitive systems and the benefits of adopting an open-systems policy for software and hardware development.

Giorgio Metta, Italian Institute of Technology and University of Genoa, inspired many of the design choices and provided leadership, guidance, and a level of commitment to the project that was crucial for its success.

Paul Fitzpatrick, Lorenzo Natale and Francesco Nori, Italian Institute of Technology and University of Genoa, together formed an indispensible team whose wide-ranging contributions to the software and hardware formed the critical core of the iCub.

José Santos-Victor and Alex Bernardino, IST Lisbon, set the benchmark early on with their design of the iCub head and their work on learning affordances, computational attention, gaze control, and hand-eye coordination.

Francesco Becchi, Telerobot S.r.l., provided the industry-strength know-how which ensured that mechanical components of the iCub were designed, fabricated, assembled, and tested to professional standards.

Rolf Pfeifer, University of Zurich, provided inspiration for the tight relationship between a system's embodiment and its cognitive behaviour, a relationship that manifests itself in several ways in the the design of the iCub.

Kerstin Rosander, University of Uppsala, was instrumental in keeping the work on developmental psychology in focus and on track.

Laila Craighero, University of Ferrara, provided many of the key insights from neurophysiology which guided our models of cognitive development.

Paolo Dario and Cecilia Laschi, Scuola S. Anna, Pisa contributed expertise in robot control systems for visual servo control and tracking.

Kerstin Dautenhahn, Chrystopher Nehaniv, University of Hertfordshire, provided guidance on social human-robot interaction.

Darwin Caldwell and Nikos Tsagarakis, Italian Institute of Technology and University of Sheffield, were responsible of the success of the design of the iCub's torso and legs.

Aude Billard and Auke Ijspeert, Ecole Polytechnique Federal de Lausanne, developed the software for imitation-based learning and gait control, respectively.

To the countless other people we haven't named — Commission officers, technicians, professors, Ph.D. and M.Sc. students, research assistants, office adminstrators, post-docs — a heart-felt thank you. This book reflects only a small part of the RobotCub project but it wouldn't have been possible without the collective effort of everyone involved in creating the iCub.

Finally, a big thank you to Keelin for her painstaking work proofreading the manuscript.

Contents

Chapter 1
A Conceptual Framework for Developmental Cognitive Systems

1.1 Introduction

This book addresses the central role played by development in cognition. We are interested in particular in applying our knowledge of development in natural cognitive systems, i.e. human infants, to the problem of creating artificial cognitive systems in the guise of humanoid robots. Thus, our subject matter is cognition, development, and humanoid robotics. These three threads are woven together to form a roadmap that when followed will enable the instantiation and development of an artificial cognitive system. However, to begin with, we must be clear what we mean by the term cognition so that, in turn, we can explain the pivotal role of development and the central relevance of humanoid embodiment.

In the following, we present a conceptual framework that identifies and explains the broad stance we take on cognitive systems — emergent embodied systems that develop cognitive skills as a result of their action in the world — and that draws out explicitly the theoretical and practical consequences of adopting this stance.[1]

We begin by considering the operational characteristics of a cognitive system, focussing on the purpose of cognition rather than debating the relative merits of competing paradigms of cognition. Of course, such a debate is important because it allows us to understand the pre-conditions for cognition so, once we have established the role cognition plays and see why it is important, we move on to elaborate on these pre-conditions. In particular, we introduce the underlying framework of enaction which we adopt as the basis for the research described in this book.

By working through the implications of the enactive approach to cognition, the central role of development in cognition becomes clear, as do several other key issues such as the crucial role played by action, the inter-dependence of perception and action, and the consequent constructivist nature of the cognitive system's knowledge.

[1] This chapter is based directly on a study of enaction as a framework for development in cognitive robotics [385]. The original paper contains additional technical details relating to enactive systems which are not strictly required here. Readers who are interested in delving more deeply into enaction are encouraged to refer to the original.

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 1–11.
springerlink.com

The framework of enaction provides the foundation for subsequent chapters which deal with the phylogeny and the ontogeny of natural cognitive systems — their initial capabilities and subsequent development — and the application of what we learn from these to the realization of an artificial cognitive system in the form of a humanoid robot.

1.2 Cognition

Cognitive systems anticipate, assimilate, and adapt. In doing so, they learn and develop [387]. Cognitive systems anticipate future events when selecting actions, they subsequently learn from what actually happens when they do act, and thereby they modify subsequent expectations and, in the process, they change how the world is perceived and what actions are possible. Cognitive systems do all of this autonomously. The adaptive, anticipatory, autonomous viewpoint reflects the position of Freeman and Núñez who, in their book *Reclaiming Cognition* [105], assert the primacy of action, intention, and emotion in cognition. In the past, however, cognition was viewed in a very different light as a symbol-processing module of mind concerned with rational planning and reasoning. Today, however, this is changing and even proponents of these early approaches now see a much tighter relationship between perception, action, and cognition (e.g. see [7, 214]).

So, if cognitive systems anticipate, assimilate, and adapt, if they develop and learn, the first question to ask is *why* do they do this? The subsequent question is *how* do they do it? The remainder of this section is devoted to the first question and the rest of the book addresses the latter.

The view of cognition taken in this book is that cognition is the process whereby an autonomous self-governing system acts effectively in the world in which it is embedded [237]. However, in natural systems, the latencies inherent in the neural processing of sense data are too great to allow effective action. This is one of the primary reasons a cognitive agent must anticipate future events: so that it can prepare the actions it may need to take. In addition, there are also limitations imposed by the environment and the cognitive system's body. To perform an action, one needs to have the relevant body part in a certain place at a certain time. In a dynamic environment that is constantly changing and with a body that takes time to move, this requires preparation and prediction. Furthermore, the world in which the agent is embedded is unconstrained and the sensory data which is available to the cognitive system is not only 'out-of-date' but it is also uncertain and incomplete. Consequently, it is not possible to encapsulate a priori all the knowledge required to deal successfully with the circumstances it will experience so that it must also be able to adapt, progressively increasing its space of possible actions as well as the time horizon of its prospective capabilities. It must do this, not as a reaction to external stimuli but as a self-generated process of proactive understanding. This process is what we mean by development. In summary, and as noted in the Preface, the position being set out in this book is that (a) cognition is the process by which an autonomous self-governing agent acts effectively in the world in which it is embedded, that (b) the dual purpose of cognition is to increase the agent's repertoire

1.6 Summary

We conclude the chapter with a summary of the principal requirements for the development of cognitive capabilities that are implied by the adoption of the enactive system stance on cognition.

A cognitive system must support two complementary processes: structural determination and development. Structural determination acts to maintain the autonomous operational identity of the system through a process of self-organizing perception and action provided by the system phylogeny. The phylogeny also provides the mechanisms for development through a process of self-modification which functions to extend the system's repertoire of possible actions and expand its anticipatory time-horizon.

The phylogeny must be capable of allowing the system to act on the world and to perceive the effects of these actions. The phylogeny must have a rich array of sensorimotor couplings and it must have a nervous system that modifies itself to facilitate the construction of an open-ended space of action-perception mappings built initially on the basis of sensorimotor associations or contingencies.

The phylogeny must allow the system to generate knowledge by learning affordances: to interpret a perception of something in its world as affording the opportunity for the system to act on it in a specific way with a specific outcome (in the sense of changing the state of the world).

The phylogeny must have some facility for internal simulation to accelerate the scaffolding of early sensorimotor knowledge and to facilitate prediction of future events, the reconstruction of observed events, and the imagination of possible new events. The phylogeny must facilitate the grounding of the simulation in action to establish its worth and thus either discard the experience or use it to enhance the system's repertoire of actions and its anticipatory capability.

Finally, the phylogeny must embody social and exploratory motives to drive development. These motives must enable the discovery of novelty and regularities in the world and the potential of the sytem's own actions.

Chapter 2
Pre-natal Development and Core Abilities

In this chapter, we consider the phylogeny of human infants and, in particular, we look at the innate capabilities of pre-natal infants and how these develop before and just after birth. We begin by looking at the role of action in cognitive behaviour, noting that anticipatory goal-directed actions, initiated by the infant in response to internal motivations, are the key to development. This is consistent with what we said in the previous chapter regarding co-development being a self-generated process. We then proceed to consider the phylogeny of a neonate and the development that occurs prior to birth. We refer to this as pre-structuring and it occurs in several guises: in the morphology of the body, in the motor system, and in the perceptual system. The resultant capabilities that exist at birth are subject to accelerated development early on. These form functional systems to sustain life and to explore and adapt to the infant's new environment. We then address the core abilities in more detail, looking at core knowledge with respect to capabilities concerning the perception of objects, numeric quantities, space, and people. This brings us to the issue of core social and explorative motives that are responsible for driving development. We conclude this chapter with a summary of the key points that enable the development of cognitive capabilities, the subject matter of the next chapter.

2.1 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

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 13–28.
springerlink.com

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 [183]. However, this is only done if an action is implied: when showing the same movements without the context of an agent, subjects fixate the moving object instead of the goal [97]. Thus, the goal state is already represented when actions are planned [185]. 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', is activated when perceiving as well as when performing an action [312]. These neurons are specific to the goal of actions and not to the mechanics of executing them [378].

Actions are guided by prospective information, as opposed to instantaneous feedback data. This is because 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 one's 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

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 [116]. 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.

2.4 Summary

We conclude with a summary of the issues addressed in this chapter. We will draw on these later in the book when we seek to identify the appropriate phylogeny for a humanoid robot, specifically a phylogeny that can support subsequent development.

2.4.1 Actions

Movements are organized as actions. The infant initiates actions as a consequence of motives, either internally-generated or externally-triggered. Actions are not reactions: actions are goal-directed and are guided by prospection. The goal state is already represented when the action is planned. For example, infants fixate the goals and sub-goals when observing actions, but if the context of a movement is removed, the fixation reverts to the motion itself rather than remaining on the anticipated outcome of the action, i.e., the goal. Similarly, the activation of mirror neurons is specific to the goal of an action and not to the movements carried out to achieve the goal. Prospective control is based both on sensory-based immediate perception and knowledge-based cognition. Prospection is possible because of the lawfulness of the world: the regularities of natural objects and the rules of social behaviour. The ultimate function of cognition is to guide actions.

2.4.2 Prenatal Development

The potential of an organism depends on the balance between phylogeny and subsequent development. Development depends on the presence of 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 cognitive abilities.

2.4.2.1 Morphological Pre-structuring

Body parts are part of a perception-action system that also includes special-purpose perceptual and neural mechanisms. Together they solve specific action problems. Consequently, changes in morphology involve matching changes in the perceptual system to improve the extraction of information for controlling specific actions. In lower vertebrates, action systems are relatively independent. A frog's perception-action system for catching flies is distinct from its perception-action system for negotiating obstacles. In higher vertebrates, movements and perceptual capabilites are more versatile and are recruited or re-used by skills other than the ones in which they evolved. The same facility for re-use also applies to subsequent ontogeny.

2.4.2.2 Pre-structuring the Motor System

Early in ontogenesis, movements are constrained to reduce the number of degrees of freedom and thereby simplify the control task. This is achieved by synergies between motor systems that both facilitate and constrain the control problem. For example, a neonate simultaneously extends its arm and fingers when reaching and, consequently, it can't grasp; the ability to grasp with the arm extended is developed later. The stepping frequency of a neonate is related to the characteristics of an the optic flow pattern observed by the neonate, another example of synergy in perception-action coupling. The key point is that there exist pre-structured sensory-motor couplings at birth, both in terms of perception-dependent movement and movement-dependent perception.

2.4.2.3 Pre-structuring the Perceptual System

Object perception — the ability to divide up the visual field into object-defining entitities — is present at birth. Early structuring of vision is accomplished prenatally in a two-stage process. First, axons orginating at the retinal level migrate to the lateral geniculate nucleus and superior colliculus. As they do so, the retinal topography is roughly preserved but not to the extent that it facilitates the extraction of useful information. Second, this mapping is refined by competitive and reinforcement interactions whereby movements of the arms in front of the eyes in the womb may facilitate the establishment of sensory-motor contingencies. After birth, visual acuity improves significantly, from 2-3% of adult acuity at birth. This resolution, however, is sufficient for an infant to see its hands and see the gross features of a person's face.

Perceptions and actions are mutually dependent: perception is needed to plan for and guide action, and action is needed to enable perception. Perception is always characterized by exploratory activities. However, the mutual dependency of perception and action is asymmetric: specific perceptual capabilities are required for certain actions but specific actions are not required to produce specific perceptions. Actions facilitate perception in a general way only: they provide opportunities for perception and guide the perceptual system to where the information is. This has an important consequence for development: the ability to extract information must exist before actions can be organized.

2.4.2.4 Forming Functional Systems

Phylogeny is geared towards sustaining life at birth. The capabilities that exist at birth are subject to accelerated development and growth early on to form and consolidate the functional systems needed to sustain life and to explore the infant's new environment and to adapt to it.

2.4.3 Core Abilities

2.4.3.1 Core Knowledge

Infants have pre-configured abilities to enable them to acquire knowledge and build on this knowledge through the developmental process. This knowledge relates to the perception of objects and their movements, the perception of geometric relationships between objects, the perception of numbers of objects, and the perception of persons and their actions. These pre-configured abilities are organized as core knowledge systems which are domain-specific (each system represents a small subset of what can be perceived), task-specific (each system solves a limited number of problems), and encapsulated (each system performs a cohesive function relatively independently of other systems).

Regarding the perception of objects, infants divide up their optic array into regions that exhibit certain characteristics. These characteristics are inner unity, a persistent outer boundary, cohesive and distinct motion, relatively constant size and shape when in motion, and, when contact occurs with another object, a change in the behaviour or motion of one or both of the objects. These characteristic regions are perceived as objects. Objects are perceived to move on continuous and regular paths which don't change if the object moves out of view, e.g., due to occlusion. When tracking an object through occlusion, an infant's gaze stops at the point of disappearance and then saccades to the point of expected reappearance *just before* the object reappears. This behaviour emerges approximately at month three and is mature by month four. However, infants are adaptive: if the expected behaviour of the object does not materialize, other expectations of a reappearance take over, e.g. expectation of reappearance at right angles to the original trajectory before being occluded.

Regarding numbers, by between six to twelve months, infants can discriminate between groups of one, two, and three objects but not higher numbers. This ability is not modality specific: it is present with hearing as well as vision. Infants can also add small numbers up to a limit of three; e.g. if you hide one object and then hide another, the infant has an expectation that two objects will be revealed. Infants can discriminate between groups of larger numbers of objects provided that the ratio of the number of each group is large, e.g. a group of eight can be discriminated from a group of sixteen, but not of twelve. In contrast, adults can successfully discriminate between a group of seven and a group of eight.

Regarding space, navigation is based on representations that are dynamic (i.e. not enduring or persistent), that are ego-centric rather than eco-centric, and that use limited amounts of information about the environment. Navigation uses path integration, navigating by moving from place to place, re-orienting as you go. Errors in navigation are due to re-orientation errors rather than errors in the recollection of landmarks. Re-orientation is effected by recognizing places or landmarks and not by using a global representation of the environment. The view-dependence of landmarks is important for re-orientation: it is the geometry of the landmark that matters rather than the distinctive features.

Regarding people, infants are attracted to people and especially to their faces, their sounds, movements, and features. Infants prefer biological motion rather than non-biological mechanical motion. Infants can recognize people and expressions and they can perceive the goal-directioned nature of actions. Infants have a greater tendency to scan a schematic face with a correct spatial arrangment of facial features rather than one where the facial features are placed randomly. Infants gaze longer when the person looks directly at them. They perceive and communicate emotions through facial gesture and they engage in turn-taking.

Core knowledge systems contribute to cognitive development in two ways. First, core knowledge systems persist in older children as domain-specific, task-specific, and encapsulated capabilities. Second, they act as building blocks for scaffolding new cognitive abilities and more complex cognitive tasks are accomplished by recruiting existing core knowledge systems in new ways.

2.4.3.2 Core Motives

The two primary motives that drive actions are social and explorative. Without social interaction a person may stop developing altogether. The social motive exists from birth and is manifested as a fixation on social stimuli (e.g. faces), imitation of basic gestures, and engagement in social interaction (e.g. in turn-taking). The explorative motive is concerned with finding out about one's own action capabilities. This motive is so strong that infants persevere in actions despite continued failure, e.g. reaching without success and abandoning a successful skill (e.g. crawling) in order to learn a new one (e.g. walking). The chief motive is to expand the space of actions.

Chapter 3
The Development of Cognitive Capabilities in Infants

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 [173, 174]. Once a critical mass of connections is established, a self-organizing process begins that results in new forms of perception, action and cognition. The emergence of new forms of action always relies on multiple developments [371]. 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.

3.1 The Development of Perception

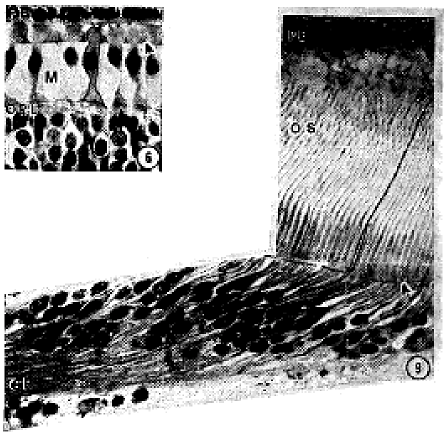
Two processes of perceptual development can be distinguished. The first one is a spontaneous perceptual learning process that has to do with the detection of structure

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 29–63.
springerlink.com

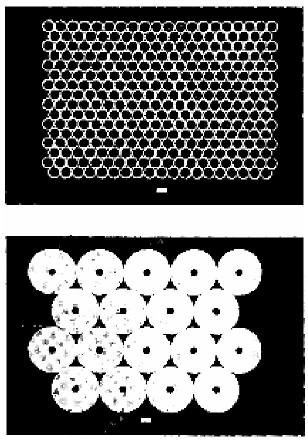
in the sensory flow. As long as there is variability and change in the sensory flow, the perceptual system will spontaneously learn to detect structure and differentiate invariants in that flow that correspond to relatively stable and predictible properties of the world. The second process is one of 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 [288, 289]. It may actually be the other way around.

3.2 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 [20] and the cones are undeveloped. Therefore both acuity and contrast sensitivity is bad. The poor visual acuity is primarly determined by the immaturity of the receptors. This is shown in Fig. 3.1. 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



(a)



(b)

Fig. 3.1 (a) Development of human foveal cones illustrated by light micrographs. Ages: (6) 5 days postpartum; (9), 72 years.; OS, outer segments (from [412]); (b) The distribution of photocells on the fovea of an adult (top) and a newborn infant (bottom). The light sensitive elements are depicted as a black spot at the center of each receptor depicted in white (from [20]).

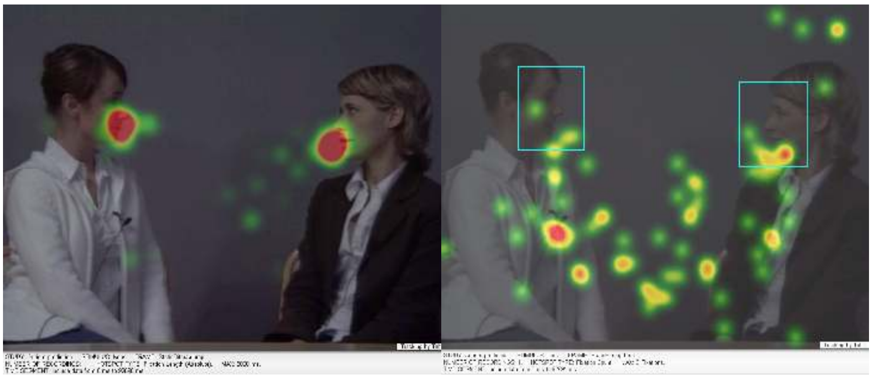


Fig. 3.15 Hotspots of looks at a conversation between two people for a typically developed 3year-old child (to the left) and for an autistic child (to the right). The intensity of looking goes from green to red where red is the most intense looking.

of age and this provides a new resource for communication. Pointing often starts when objects are named, an example that language and planned directed actions are connected [376].

While young children extremely attracted by other peoples faces and spend much time looking at them, children with autism dont. Their attention is drawn to simpler features that are salient, like high contrast or bright colour. Figure 3.15 shows a typically developing child and a child with autism who looks at two people having a conversation. While the typically developing child focuses on the mouths of the talking people, the child with autism does this much less and devotes a fair amount of time looking at the shadow in between.

3.4 Summary

We conclude with a short synthesis of the many issues addressed in this chapter. In doing so, we will highlight the key points and, where relevant, provide the timeline for development of certain abilities.

3.4.1 The Basis for Development

Development arises due to changes in the central nervous system as a result of dynamic interaction with the environment. Development is manifested by the emergence of new forms of action and the acquisition of predictive control of these actions. Mastery of action relies critically on prospection, i.e. the perception and knowledge of upcoming events. 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.

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. There are at least two exploratory motives, one involving the discovery of novelty and regularities in the world and one involving the discovery of the potential of the infant'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). Similarly, the discovery of what objects and events afford in the context of new actions is also a strong motivation.

The emergence of new forms of action always relies on multiple developments, typically in perception and prospective motor control. In the development of perception, there are two processes: the detection of structure or regularity in the flow of sensory data, and the selection of information which is relevant for guiding action.

3.4.2 Visual Processing

The visual system develops rapidly after birth. The acuity and contrast sensitivity of the retina in a newborn infant is very poor, typically 2.5% – 3.5% of their eventual sensitivity. Both develop very quickly and the acuity of an adult is achieved by month 5. This development occurs through the migration of cones to the fovea; rods don't change position. This migration changes the structure of the retina and it means that infants do not have innate sensitivity of certain retinal patterns which must be learned after birth.

Several other visual functions are not available at birth. Colour perception functions only after approximately four weeks. Motion perception exists at a sub-cortical level but the interpretation and use of motion perception requires cortical processing and does not exist at birth. For example, neonates cannot do smooth pursuit at birth and only begin to improve at week 6, achieving adult performance at week 14. The ability to discriminate between regions of different directions of motion also only emerges at approximately week 8 and is mature by weeks 14-18.

Visual space perception depends of several cues all of which develop in the first year and at different rates. These cues include binocular depth perception based on vergence of the eyes and stereo disparity. Vergence develops from week 4 for distances greater than 20 cm and by week 20 or perhaps earlier it can be used for reaching actions. Sensitivity to stereo binocular disparity develops quickly from week 8 on. Depth perception due to motion parallax caused by movements of the infant's head develops by approximately week 12. The ability to estimate the time to contact of a looming object is not perceived by very young infants and is only apparent from month 6 on. Infants younger than six months may know an object is approaching but they are unable to tell when it is going to hit them. Other depth cues such as perspective, size, interposition, and shading are not used to guide reaching until months 6–7.

Young infants primarily identify objects using binocular stereo disparity and relative motion. Objects are perceived as entities with well-defined outer boundaries and inner uniformity. Relative motion dominates the perception of objects in very young infants and is much more important than static features such as form and similarity or uniformity of texture and colour. That is to say, Gestalt relations influence infants far less than they do adults.

Infants divide the perceptual array into entitites that move together or move separately, that maintain size and shape during motion, and that tend to act on each other only when they make contact. Entities that move together are perceived as a single object, even if they comprise disconnected regions, so that different parts of a partially-occluded object are perceived as a single object, provided there is relative movement between the object and the occluder due either to independent object motion or motion parallax. This is true irrespective of the similarity of the disconnected regions. Entities that move relative to one another are perceived as separate objects. The alignment of the disconnected regions does have an impact, with the perception of a single object being stronger if there is good alignment. Colour contributes to the identification of an object only by the end of year 1. The combination of visual and tactile exploration of an object in infants of approx. 11 months increases the ability to distinguish objects based on colour.

The following shows the timeline for the onset of development of visual processing.

Month	
0	Visual acuity
1	Colour processing
1	Ocular convergence for objects beyond 20 cm
2	Depth perception from binocular stereo disparity
2	Object discrimination based on motion information
3	Smooth pursuit
3	Depth perception from motion parallax
3	Ability to perceive binocular depth
5	Ocular convergence for reaching
6	Depth perception of looming objects
6–7	Depth perception based on perspective, size, interposition, shading
12	Object discrimination based on colour

3.4.3 Posture

Establishing and maintaining a stable orientation with respect to the environment is a pre-requisite for purposeful movements, i.e. actions. Gravity provides a frame of reference and is sensed using the vestibular system (in particular, using the otoliths in the ear). Vision, and visual flow in particular, is crucial for maintaining balance and controlling body posture prospectively. It is important to maintain balance prospectively because reflexive posture adjustment interrupts actions. For the same reason, the effect of the movement of the limbs on balance is also adjusted prospectively.

The first signs of being able to control posture begins at week 12 as the infants stabilizes head pose while lifting the head when prone. By weeks 24–28, the infant can stabilize head posture when sitting, compensating for sway in the trunk. By week 36, the infant is making anticipatory adjustments of head and trunk posture when reaching.

The development of locomotion provides several instances of the interdependency of perception and action. For example, infants who can walk show a preference for looking at other infants walking whereas infants who crawl show a preference for looking at other infants crawling. These examples suggest that a mirror-neuron function is involved.

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

Month	
3	Head stabilization when lying prone and lifting the head
6	Head stabilization when sitting
9	Anticipatory adjustment of posture when reaching

3.4.4 Gaze

Crucial for the establishment and maintenance of social interaction, the development of gaze control is one of the earliest skills to appear in a neonate. Gaze involves both head and eye movements and is guided by visual, vestibular, and proprioceptive information. To develop prospective gaze control, the infant must master two skills: high-speed saccadic movements to areas of interest with subsequent gaze stabilizations, and smooth pursuit eye movements.

Controlling saccadic movements is a basic aspect of cognitive development. Shifting gaze is preceded by a covert attentional shift: a disengaging of attention on a current point of interest, engagement on a new point of interest, followed by an overt shift in gaze. The saccadic system develops ahead of the smooth pursuit system. It is functional at birth and develops rapidly in the first six months. Visual attention in infants is primarily guided by the attractiveness of objects and the predictability of events. The systematic scanning of the environment only appears at pre-school age at which point a child can solve the problem of detecting the difference between two pictures.

Smooth pursuit is more complicated that saccadic movement as it requires anticipation of imminent motion of the object of interest. Smooth pursuit is also needed when the infant is moving with respect to the object of interest and here the stabilization needs to anticipate body movements. Newborn infants have only a limited ability to track moving objects smoothly but they improve rapidly from around week 6 and attain adult level by week 14. When tracking a moving object, the smooth pursuit system must anticipate motion. One predictive process extrapolates just-observed motion, i.e. instantaneous motion. Such extrapolation depends on the regularity of motion due to, e.g., inertia. Both head and eye movements are involved in tracking, with the eyes leading the head and the head often lagging the target by 0.3 sec. (for a 5-month-old infant). The head tracking may actually interfere with the eye-tracking at this age but infants persist because they are internally-motivated to do so (and because, like the transition from crawling to walking, the combined head-eye tracking eventually develops into a much more flexible skill).

Head movements also occur for reasons other than gaze stabilization. Since the visual system, compensating for retinal slip, works best with slow changes in the optical field less than 0.6 Hz, it is the vestibular system, operating best at frequencies greater than 1 Hz, that is used to stabilize gaze and compensate for head movement.

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

Month	
0	Vestibular gaze stabilization to compensate for head movement
0	Saccadic eye movements, ability to engage and disengage attention
0	Limited smooth pursuit ability
0	Attentional processes are present: gaze directed toward attractive objects and novel appearance or events
3–4	Infants achieve adult level of smooth pursuit

3.4.5 Reaching and Grasping

Visual control of the arm is present at birth. Infants can move their fingers but cannot control them to grasp or manipulate objects. Arm and finger motions are bound together in extension and flexion synergies, i.e., the arm and fingers extend and flex together. This synergy changes dramatically at approximately week 8 and the coupling is broken. Now the infant has a tendency to fist the hand when extending the arm. This is followed after a few weeks with open-handed reaching, but only when the arm is visually guided to an object with the hand closing when it is close to the object. Reaching for a stationary object appears between weeks 12 and 18. Catching moving objects appears at approximately week 18, i.e. the age at which an infant masters reaching for stationary objects. Significantly, the point towards which the infant reaches is the eventual position of the moving object, not the initial observed position. Again, we see the presence of prospection in an infant's actions. Early reaching movements are characterized by several segments or units, each comprising an acceleration and a deceleration phase. The number of units reduces with development to the point where infants of a few months of age making movements comprising two segments: a reaching movement to bring the hand close to the target and a subsequent grasping movement. The development of reaching abilities shows increasing use of prospective control. Six-month-old infants do not plan reaching actions when the target object is temporarily occluded; instead they wait until the object re-appears. Nine-month-old infants reach for moving objects but inhibit the reaching action when a barrier blocks their path. Eleven-month-old infants are able to catch an object as it reappears from behind an occluder, exhibiting significant anticipatory control.

Infants also use prospection when preparing the grasping action. They adjust the orientation of the hand to the orientation of the object they are reaching for. In particular, the hand orientation is adjusted to align with the future anticipated orientation of the target object. Between 9 and 13 months, infants also adjust the extent of the hand opening to match the size of the object to be grasped; infants do not exhibit this behaviour at five months old. All infants begin to close the grasp before the object is encountered, again showing the operation of prospection in actions. The exact behaviour differs with age. Infants up to 9 months old first move the hand close to the object and then begin the grasping action. Thirteen-month-olds begin the grasp during the approach and significantly before the hand touches the objects. Eventually, the infant displays an integrated reach and grasp action.

When grasping first emerges, infants may use one as well as both hands. The first grasps are power grasps which engage the whole hand. At month 4, infants adjust grasp configurations as a function of the object properties. However, in doing so, they use both haptic and visual information to adjust the grip. By month 8, most infants are able to do this using only visual information. From months 5 to 9, infants increasingly use differentiated grip positions involving the thumb, index finger, and long middle finger. From months 9 to 10, infants develop independent finger control in grasping allowing them to grip very small objects in a precision grasp. When adjusting the grasp to the orientation of an elongated object, the precision of the adjustment increases with age but a residual error of approx. 10° – 15° remains which must be accommodated after contact. Again, grasping is prospective with infants adjusting for the final orientation of the object.

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

Month	
0	Visual control of arm but no control of fingers for grasping
0	Arm and finger motions governed by global extension/flexion synergies
2	Hand is fisted when extending the arm
3	Open hand when reaching, but only when visually guided hand closing when close to object
4–5	Reaching and grasping as a function of object properties
9	Adjustment of hand size when reaching hand closes when in vicinity of object
9–10	Differentiated finger grasping, e.g. pincer grasp
13	Grasping starts when reaching: i.e. one integrated reach-grasp act

3.4.6 Manipulation

Even when only one hand is used for grasping object, two hands are most often used to manipulate it, typically to inspect it from several viewpoints. When manipulating objects, the infant has to imagine the goal state of the manipulation and the strategy by which the goal state can be achieved. For tasks such as inserting an object into a matching aperture, the infant does not develop this ability until at least month 22. Again, prospection and internal simulation to imagine a goal status and perform mental rotations of a manipulated object are required. A pure feedback strategy does not work for this task.

3.4.7 Social Abilities

Newborn infants understand basic emotions communicated by facial gestures and they imitate these gestures from birth when engaging in face-to-face interactions. Newborns can perceive the direction of attention of others and by month 10 – 12 they can reliably follow gaze. Social interaction relies primarily on vision, touch, and proprioception using the mouth, face, hands, and eyes. Since the infant is interacting with other cognitive agents rather then physical objects obeying physical laws, infants must learn the conventions of social interaction, intention, and emotion in order to engage the necessary prospective skills required for effective action. Intention and emotions are conveyed by elaborate and specific movements, gestures, and sounds and neonates are very attracted to the sounds, movements, and features of the human face.

Chapter 4
What Neurophysiology Teaches Us About Perception and Action

We now shift from developmental psychology to neurophysiology to focus on the relationship between perception and action. In particular, we are interested in discovering what we can learn about the way a primate brain handles the perception of space, the perception of objects upon which the primate can act, structured interaction, and selective visual attention. In doing so, we will be concerned in particular with teasing out the dependency of perception on actions, both actual and potential. What we learn from this exercise results from a shift in our understanding of the way different parts of the brain interoperate. This shift represents a move away from a prevalent view of a complete separation of function among the dorsal and ventral streams in the brain, the former supposedly dealing exclusively with issues of location and space, the latter supposedly dealing exclusively with issues of identity and meaning. Instead, what emerges is a picture in which the dorsal stream plays a very active role in the recognition of actions and in object discrimination due to their affordances. We will also see that perceptions are directly facilitated by the current state of the premotor cortex.

4.1 The Premotor Cortex of Primates Encodes Actions and Not Movements

In its neurophysiological sense, the term "action" defines a movement made in order to achieve a goal. The goal, therefore, is the fundamental property of the action. There are actions aiming to reach and manipulate objects, actions aimed towards oneself, and communicative actions. The traditional approach to the cortical motor system has always focused on the study of movement and not on that of action. This is mainly because the electrical stimulation of the primary motor cortex, the more excitable one, evokes movements. Consequently, Frisch and Hitzig at the end of the nineteenth century and, subsequently, Ferrier, interpreted the results of the electrical stimulation of the motor cortex of the dog and monkey as proof of the existence of a map of movements in the cortex. Furthermore, around the middle of the twentieth century, Penfield in humans [284] and Woolsey in macaques [406], using surface

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 65–79.
springerlink.com

electrical stimulation studies during neurosurgical operations on humans and experiments on monkeys respectively, defined high-resolution somatotopic motor maps. Even today, no neuroscience text is complete without showing these two homunculi with enormous hands and mouths, one standing on its head on the prerolandic cortex (the larger, MI) and the other lying on the mesial frontal cortex (the smaller, SMA).

Despite the fact that these maps can be didactically useful to demonstrate basic and essential concepts, such as that of somatotopic arrangement (adjacent points of the body are represented in adjacent points of the cortex) and anisotropy (the extension of the cortical motor representation of a body part is proportional to the complexity of the movements represented therein and not to the physical dimension of the body part), Penfield's homunculus and Woolsey's figurine consolidated the pervasive dogma of clinical neurophysiology according to which movements are represented in the cortex. Moreover, according to this view, the motor system comprises just two areas: the primary motor area (MI) and the supplementary motor area (MII or SMA).

Neuroanatomical evidence was used to support this functional representation of the motor cortex. In fact, the frontal motor cortex has an agranular and fairly homogenous cytoarchitectonic structure, with no dramatic differences between sectors. For this reason, in 1909, basing his conclusions on the distribution of pyramidal cells, the famous German neuroanatomist, Korbinian Brodmann [48], suggested that the motor cortex of primates was formed by two areas (area 4 and area 6) which, in their extension, almost totally comprised Penfield's and Wolsey's cartoons. In spite of early criticism of the inadequacy of this division by some of Brodmann's colleagues (including the Vogt partners in 1919 [388]), the existence of one motor area (area 4, considered as the main origin of corticospinal projections) and one premotor area (area 6, considered as responsible for the preparation of movements) was accepted by neurophysiologists as a good representation of current knowledge.

The vast initial consensus for this anatomical and functional reference framework can be explained by the fact that it gave a simple explanation to a complex problem: the motor system was seen as a map of movements (a sort of look-up table), connected at the output to the spinal cord servomechanism and at the input to the so-called "associative areas", responsible for integrating the various pieces of sensorial information (visual, auditory, tactile, etc.) with those inside the system (intentionality, motivation, memory) to generate motor programs. In the last two decades, this unitary vision has been challenged by a number of experimental observations:

1. Neuroanatomical studies, in particular those concerning the cortical cytoarchitecture (the study of the arrangement of the various cell types within the layers forming the cortex), histochemistry (the study of the cellular biochemical properties differentiating the various areas of the cortex) and neurochemistry (the study of the neurotransmitter and neuromodulator receptors on the cell membrane of the neurons), indicate that the agranular/disgranular cortex of the frontal lobes (this definition refers to the cellular distribution in the fourth cortical layer, generally not well represented in motor areas) is formed by a constellation of separate areas, of which the primary motor cortex (MI or Brodmann's area 4) occupies just the rearmost part, being almost entirely buried inside the central

4.6 Structured Interactions

Paul Broca established that the Inferior Frontal Gyrus (IFG) is critically important for the perception of speech. Subsequently, it has been shown that Broca's area is also involved in speech production [82, 83], especially syntactically-complex and/or ambiguous material [94]. Recent research results indicate that Broca's area is not limited to speech production and comprehension but that it may also play a part in the observation and execution of action, and in music execution and listening. What is significant about these three areas — speech, action, and music — is that all involve complex hierarchical dependencies between constituent elements: words, movements, and sounds, respectively. Significantly, there is little or no activation in Broca's area when subjects observe meaningless gestures: it is the observation of actions as goal-directed motivated intentional sequences of motor acts that triggers activation [131]. It has been argued that Broca's area represents the syntactic rules of these actions, rather a than simple basic motor program to execute the constituent movements [131] and that, in general, Broca's area might be the centre of a brain network that can encode hierarchical structures regardless of their use in action, language, or music [84]. That is, Broca's area might realize a supramodal or polymodal sensorimotor representation of hierarchical syntactic structures: the brain appears to provide an innate cortical circuit that is capable of developing to learn complex hierarchical representations of regularities that are then deployed to produce and perceive complex structured interactions [90]. Again, it is important to emphasize that Broca's area is associated with intentional events and not just simple sequences of physical states: what is crucial is the hierarchical complexity of the pattern and the motivated goal-orientation of the event. This suggest process of inference of the underlying intentionality echoing strongly the anticipatory nature of cognition.

4.7 Summary

The classic flow-diagram describing how sensory information is serially processed, and eventually transformed into movements by the brain, has become more and more implausible because of neuroanatomical and neurophysiological evidence. In particular:

1. Cytoarchitectonical, histochemical and neurochemical studies indicate that the motor cortex is indeed formed by a constellation of distinct areas, each one bidirectionally connected with a specific area of the parietal lobe.
2. The neurophysiological study of these parieto-frontal connections suggests that they might play a crucial role in effector-specific sensorimotor transformations.
3. Several motor neurons discharge also during sensory stimulation.

Accordingly, visual stimulation modulates the activity in LIP-FEF neurons, objects entering the peripersonal space activate F4-VIP neurons, graspable objects and actions of other individuals visually activate 'canonical' and 'mirror' visuomotor neurons belonging to the F5-AIP-PFG fronto-parietal circuit. This association stimulus-response, present at single neuron level, might provide the goal to the movement, thus transforming the latter into an action and, perhaps more interestingly, it might provide the basis for an attentional system which modulates, by predictive mechanisms, our understanding of the environment surrounding us.

4.7.1 Grasping

The ventral premotor cortex receives strong visual inputs from the inferior parietal lobule. These inputs subserve a series visuomotor transformations for reaching (area F4) and grasping (area F5). Area F5 is located in the rostral part of the ventral premotor cortex where intracortical microstimulation reveals extensively overlapping representations of hand and mouth actions. Single neuron studies have shown that most F5 neurons code for specific actions, rather than the single movements that form them. It has been therefore proposed that, in area F5, a vocabulary of goals more than a set of individual movements, is stored. This goal-directed encoding, typical of area F5, is demonstrated by the discriminative behavior of F5 neurons when an action, motorically similar to the one effective in triggering neuron response, is executed in a different context. The motor responses of the F5 neurons vary in their degree of abstraction, from the general encoding of an action goal (e.g., grasping, holding) to more specific responses related to particular aspects of the same goal (e.g., precision grip, whole hand grasping). Finally, there are neurons responding to different phases of these actions (e.g., during opening or closing the fingers while executing a specific grasping). Several F5 neurons, in addition to their motor properties, respond also to visual stimuli. According to their visual responses, two classes of visuomotor neurons can be distinguished within area F5: canonical neurons and mirror neurons [313]. The canonical neurons are mainly found in that portion of area F5 which is the main target of parietal projections coming from area AIP. These neurons respond to visual presentation of three-dimensional objects. About one quarter of F5 neurons show object-related visual responses, which are, in the majority of cases, selective for objects of certain size, shape and orientation and congruent with the motor specificity of these neurons. They are thought to take part in a sensorimotor transformation process dedicated to the selection of the goal-directed action which most properly fits to the particular physical characteristics of the object to be grasped.

4.7.2 Spatial Perception

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 [379], and that the posterior parietal cortex acts as a unique site for space perception. However, recent evidence suggests 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 [316, 314]. 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 reference 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). That is, the FEF neurons encode stimulus position in a retinotopic frame of reference and the LIP neurons encode the eye direction; together the LIP-FEF circuit yields a perceptuo-motor encoding of space in a craniocentric frame of reference. Similarly, movement-related neurons and sensorimotor neurons in the VIP-F4 circuit are activated by head movements, face movements, or arm movements while sensory and sensorimotor neurons respond to tactile or to tactile and visual stimuli but the visual RFs of these neurons are anchored to the tactile ones regardless of eye position. That is, the VIP-F4 circuit yields a perceptuo-motor encoding of space in a peripersonal frame of reference.

4.7.3 Perception-Action Dependency

Not only is spatial information derived and encoded in action-specific mechanisms, there is also evidence that the distinction between perception for action control and perception for semantic understanding is not valid. On the contrary, it appears that the motor system is very much involved in the semantic understanding of percepts with procedural motor knowledge and internal action simulation being used to discriminate between 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 [109]. In addition, the act of imagining (or visualizing) a grasping action [78, 122] or discriminating between hand shapes [281] also involves the dorsal stream and the premotor and inferior parietal areas.

4.7.4 Structured Interactions

Broca's area is arguably the centre of a brain network that can encode hierarchical supramodal or polymodal sensorimotor representations of hierarchical syntactic structures [84]. Thus, the brain appears to provide an innate cortical circuit that is capable of developing to learn complex hierarchical representations of regularities that are then deployed to produce and perceive complex structured interactions [90]. These interactions are intentional events and not just simple sequences of physical states so that these representations are used to encode and infer the underlying intentionality, thereby manifesting the anticipatory nature of cognition.

4.7.5 Selective Attention

Selective attention is not a unitary system as is often thought but rather it is a process that derives from the several cortical circuits and subcortical centres that are involved in the transformation of spatial information into actions. This premotor theory of attention holds that attention derives from the mechanisms that generate action. The preparation of a motor program in readiness to act in some spatial regions predisposes the perceptual system to process stimuli coming from that region.

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 [71]. Similarly, if something is within reach, attention to it is enhanced; if not, attention is diminished.

The premotor theory of attention applies not only to spatial attention but also to selective attention in which some objects rather than others are more apparent. For example, the preparation of a grasping action predisposes attention to objects that match the grasp configuration [70]. As we have seen several times before, a subject's actions conditions its perceptions.

Chapter 5
Computational Models of Cognition

Having looked at the development of cognitive abilities from the perspective of psychology and neuroscience, we now shift our sights to recent attempts at building artificial cognitive systems. In the following, we will focus on cognitive architectures rather than on specific cognitive systems. We do this because cognitive architectures are normally taken as the point of departure for the construction of a cognitive system and they encapsulate the various assumptions we make when designing a cognitive system. Consequently, there are many different types of cognitive architecture. To provide a framework for our survey of cognitive architectures, we must first address these assumptions and we will do this by beginning our discussion with an overview of the different paradigms of cognition. We have already discussed one approach to cognition in Chap. 1 — enaction — and we will take the opportunity here to position enaction within the overall scheme of cognition paradigms. With the differences between the different paradigms of cognition established, we can then move on to discuss the importance of cognitive architectures and survey the cognitive architecture literature, classifying each architecture according to its paradigm and highlighting the extent to which each architecture addresses the different characteristics of cognition. We will make this survey as complete as possible[1] but we will emphasize those architectures that are intended to be used with physical robots and those that provide some developmental capability. On the basis of this review of cognitive architectures — both general characteristics and specific instances — we will conclude with a summary of the essential and desirable features that a cognitive architecture should exhibit if it is to be capable of forming the basis of a system that can autonomously develop cognitive abilities.

5.1 The Three Paradigms of Cognition

Broadly speaking, there are three distinct approaches to cognition: the *cognitivist* approach, the *emergent systems* approach, and *hybrid* approaches [63, 382, 387] (see Fig. 5.1).

[1] The survey of cognitive architectures is an updated and extended version of a survey that appeared in [387].

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 81–99.
springerlink.com

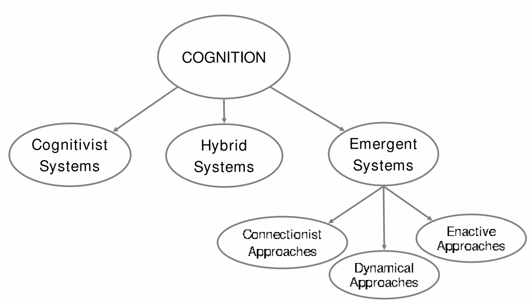


Fig. 5.1 The cognitivist, emergent, and hybrid paradigms of cognition

Cognitivist approaches correspond to the classical view that cognition is a form of symbolic computation [298]. Emergent systems approaches view cognition as a form of self-organization [195, 370]. Emergent systems embrace connectionist systems, dynamical systems, and enactive systems. Hybrid approaches attempt to blend something from each of the connectionist and emergent paradigms. Although cognitivist and emergent approaches are often contrasted purely on the basis of symbolic computation, the differences are much deeper [387, 384]. We can contrast the cognitivist and emergent paradigms on fourteen distinct characteristics:[2]

1. computational operation,
2. representational framework,
3. semantic grounding,
4. temporal constraints,
5. inter-agent epistemology,
6. embodiment,
7. perception,
8. action,
9. anticipation,
10. adaptation,
11. motivation,
12. autonomy,
13. cognition,
14. philosophical foundation.

Let us look briefly at each of these in turn (see Table 5.1 for a synopsis of the key issues).

Computational Operation

Cognitivist systems use rule-based manipulation of symbol tokens, typically but not necessarily in a sequential manner.

[2] These fourteen characteristics are based on the twelve proposed by [387] and augmented here by adding two more: the role of cognition and the underlying philosophy. The subsequent discussion is also an extended adaptation of the commentary in [387].

different ways, and differently encoded in ways that are particular to different effectors. On the other hand, the SASE cognitive architecture incorporates explicit self-modification by monitoring and altering its own state, specifically to generate models and predict the outcome of actions. In essence, this is a sophisticated process of homeostasis, or self-regulation, which preserves the autonomy of the system while allowing it to operate effectively in its environment. The Cognitive-Affective cognitive architecture schema adopts a similar approach, but extends it by proposing a spectrum of homeostasis. Different levels of cognitive function and behavioural complexity are brought about by different levels of emotion, ranging from reflexes, through drives, to emotions and feelings, each of which is linked to a homeostatic autonomy-preserving self-maintenance process, ranging from basic metabolic processes, through reactive sensorimotor activity, associative learning and prediction, to interoception and internal simulation of behaviour prior to action.

The hybrid cognitive architectures fall somewhere in between the cognitivist and emergent cognitive architectures in the extent to which they address the seven characteristics. LIDA [17, 103, 104, 106, 299] (see Sect. A.3.5) CLARION [362, 363, 364] (see Sect. A.3.6) and PACO-PLUS [201] (see Sect. A.3.7) are the only hybrid cognitive architectures that address adaptation in the developmental sense. LIDA addresses anticipation and adaptation by deploying transient and consolidated episodic memories of past experiences and procedural memory of associated actions and outcomes. Adaptation occurs when sensory-derived perceptions re-combine with associated recalled episodic memories and are either incorporated into the episodic memory as a new experience or are used to reinforce existing experiences. Selected episodic memories are used to recall associated actions and likely outcomes from the procedural memory for subsequent execution. Both anticipation and adaptation are effected in CLARION by observing the outcome of a selected action and updating bottom-level reinforcement learning in a connectionist representation and a top-level rule-based update in a symbolic representation. In addition, rule-based actions are either generalized or refined, depending on the outcome. In a similar way, the PACO-PLUS cognitive architecture also learns from observations. In particular, PACO-PLUS autonomously learns co-joint object-action affordances by exploration, selecting an action and observing the effects of these actions on the objects in the robots environment.

5.3 Summary

We conclude the chapter with a summary of the principal requirements for the design of a cognitive architecture which adheres to the emergent paradigm of cognition and, in particular, to the enactive systems approach. Such a cognitive architecture must be capable of supporting the development of cognitive capabilities, as well as adequately addressing the seven characteristics of embodiment, perception, action, anticipation, adaptation, motivation, and autonomy discussed in the previous sections. These requirements derive from a consideration of both the generally-desirable characteristics of cognitive architectures discussed in Sect. 5.2.3 and the features of specific cognitive architectures surveyed in Sect. 5.2.4.

A cognitive architecture should have the following characteristics.

1. A minimal set of innate behaviours for exploration and survival (i.e. preservation of autonomy) [203, 206, 204].
2. A value system — a set of task non-specific motivations — that guides or governs development [203, 206, 204].
3. An attentional mechanism [217].
4. Learning from experience the motor skills associated with actions [217].
5. A spectrum of self-regulating autonomy-preserving homeostatic processes (ranging from basic metabolic processes, through associative learning and prediction, to interoception and internal simulation) associated with different levels of emotion (ranging from reflexes, through drives, to emotions and feelings) resulting in different levels of cognitive function and behavioural complexity [264, 415] (Cognitive-Affective) and [394, 395, 393] (SASE).
6. Anticipation and planning based on internal simulation of interaction with the environment [335, 336, 337, 338] (Global Workspace).
7. Action selection modulated by affective motivation mechanisms [335, 336, 337, 338] (Global Workspace).
8. Separate and limited representations of the world and the task at hand in each component / sub-system [170, 171] (Cerebus)
9. Transient and consolidated (generalized) episodic memories of past experiences [17, 103, 104, 106, 299] (LIDA).
10. Procedural memory of actions and outcomes associated with episodic memories [17, 103, 104, 106, 299] (LIDA).
11. Learning based on comparison of expected and observed outcomes of selected actions, resulting in either generalization or refinement of the associated action [362, 363, 364] (CLARION).
12. Learning co-joint object-action affordances by exploration [201] (PACO-PLUS).
13. Hierarchically-structured representations for the acquisition, decomposition, and execution of action-sequnce skills [60, 213, 214, 215, 216](ICARUS).
14. Concurrent operation of the components / sub-systems of the cognitive architecture so that the resultant behaviour emerges as a sequence of states arising from their interaction as they compete and co-operate [335, 336, 337, 338] (Global Workspace).

In the next chapter, we turn our attention to integrating these requirements with those derived from our study of psychology and neurophysiology in previous chapters.

Chapter 6
A Research Roadmap

In Chapter 1 we discussed the principles of developmental cognitive systems in general, and of enactive systems in particular. Chapters 2, 3, and 4 identified the constraints arising from the developmental psychology and neurophysiology of neonates, while Chap. 5 revealed a number of insights derived from several computational models of cognition. Now we weave all of these constraints, requirements, and insights together to produce a comprehensive list of functional, organizational, and developmental guidelines for an artificial system that is capable of developing cognitive abilities. These guidelines provide the basis for the design of an enactive cognitive architecture and its practical deployment. In other words, they define a roadmap for the development of cognitive abilities in a humanoid robot, a roadmap which embraces both phylogeny and ontogeny. In the next chapter, we describe the current status of a project to implement these guidelines in a cognitive architecture for the iCub humanoid robot. This cognitive architecture, together with the physical robot, provides the platform for the development of cognitive abilities. The developmental process — or ontogenesis — must proceed in a structured manner. Consequently, we will draw heavily on the material in Chap. 3 on the development of human infants to inform this structure and present a roadmap for ontogenesis. Thus, our roadmap has two sides: the phylogenetic side, informed by enaction, developmental psychology, neurophysiology, and computational modelling, and the ontogenetic side, informed by developmental psychology (see Fig. 6.1). We begin by addressing the phylogeny of the system in Sect. 6.1 and then turn to its ontogeny in Sect. 6.2.

Before proceeding, a note on research roadmaps is in order. Arguably, the term *roadmap* has become somewhat debased in recent years due to it frequently being used to define over-ambitious research agenda. Nonetheless, a properly-constituted research roadmap has an important role to play in advancing challenging new disciplines, such as artificial cognitive systems. As Sloman has pointed out, understanding the requirements of cognitive systems research is in itself a major research activity and a research roadmap provides a way of expressing an agreed specification of what the problems are and it helps research planning by identifying milestones

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 101–119.
springerlink.com

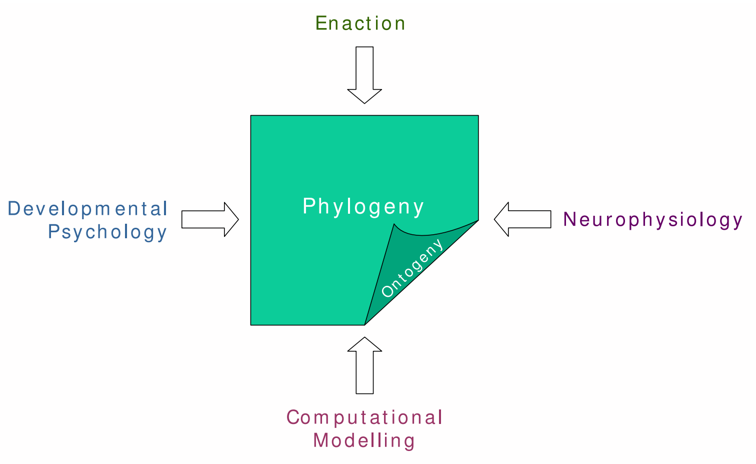


Fig. 6.1 Four influences contributing to the phylogentic and ontogenetic principles of developmental cognitive systems

and routes through them [348]. One of Sloman's main points is that "even people who disagree on mechanisms, architectures, representations, etc. may be able to agree on requirements". He argues for the collection of many possible scenarios based on observation of "feats of humans (e.g. young children, playing, exploring, communicating, solving problems) and other animals (e.g. nest-building birds, tool makers and users, berry-pickers and hunters)". These scenarios should be described in detail and then analysed in depth to produce requirements which, together with the scenarios, should be ordered by difficulty and by dependence. Sloman conjectures that

> "the most general capabilities of humans, which are those provided by evolution, and which support all others, develop during the first few years of infancy and childhood. We need to understand those in order to understand and replicate the more 'sophisticated' and specialised adults that develop out of them. Attempting to model the adult competences directly will often produce highly specialised, unextendable, and probably very fragile systems – because they lack the child's general ability to accommodate, adjust, and creatively re-combine old competences" [348].

While not following the exact methodology for creating a research roadmap advocated by Sloman, the roadmap described in this book nonetheless adheres strongly to this philosophy, basing the requirements encapsulated in the guidelines and scenarios below on cognitive development in human infants.

Chapter 7
The iCub Cognitive Architecture

In this chapter, we examine how the roadmap guidelines set out in Chap. 6 have influenced the design of a cognitive architecture for the iCub[1] humanoid robot. We begin with a overview of the iCub and we discuss briefly the iCub mechatronics and software infrastructure.[2] We then describe the iCub cognitive architecture, focussing the selection of a minimal set of phylogenetic capabilities derived from the seven groups of roadmap guidelines. Since the iCub cognitive architecture is a work-in-progress, it represents only a partial implementation of the roadmap guidelines. Consequently, we close the chapter by examining the exact extent to which each guideline has been followed. The next chapter, which concludes the book, addresses some of the research challenges posed by a complete implementation of the roadmap guidelines.

7.1 The iCub Humanoid Robot

The iCub is an open-systems 53 degree-of-freedom humanoid robot [249, 330]. It is approximately the same size as a three or four year-old child (see Fig. 7.1) and it can crawl on all fours and sit up. Its hands allow dexterous manipulation and its head and eyes are fully articulated. It has visual, vestibular, auditory, and haptic sensory capabilities. The iCub was designed specifically as an open-systems research platform for the embodied cognitive systems community. It was conceived with the aim of fostering collaboration among researchers and lowering the barrier to entry into what for many would otherwise be a prohibitively expensive field of research. The iCub is licensed under the GNU General Public Licence (GPL)[3] so

[1] iCub stands for Integrated Cognitive Universal Body and was motivated by the 'I' in Asimov's *I, Robot* and 'cub' from Mowgli the man-cub in Kipling's *Jungle Book*.
[2] Sect. 7.1 is based directly on a description of the iCub in [249].
[3] The iCub software and hardware are licensed under the GNU General Public Licence (GPL) and GNU Free Documentation Licence (FDL), respectively.

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 121–153.
springerlink.com

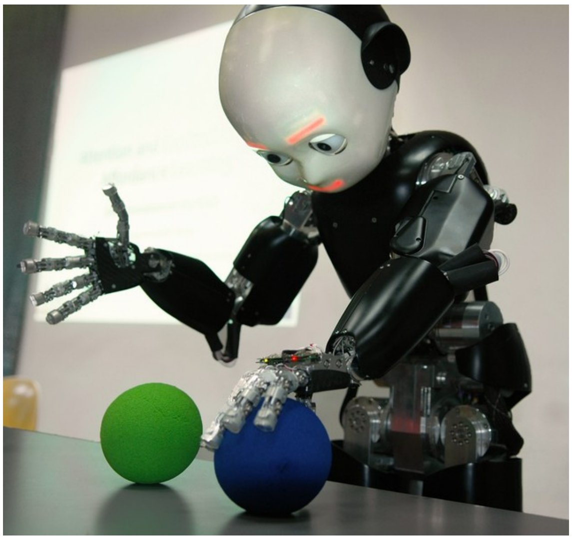


Fig. 7.1 The iCub humanoid robot: an open-systems platform for research in cognitive development

that researchers can use it and customize it freely. To date, twenty iCubs have been delivered to several research labs in Europe and to one in the U.S.A.[4]

7.1.1 The iCub Mechatronics

The iCub is approximately 1m tall and weighs 22kg. From kinematic and dynamic analysis of required motion, the total number of degrees of freedom (DOF) for the upper body was set to 38 (7 for each arm, 9 for each hand, and 6 for the head). For each hand, eight DOF out of a total of nine are allocated to the first three fingers, allowing considerable dexterity (see Fig. 7.2). The remaining two fingers provide additional support for grasping. Joint angles are sensed using a custom-designed Hall-effect magnet pair. In addition, tactile sensors are under development [228]. The overall size of the palm has been restricted to 50mm in length; it is 34mm wide at the wrist and 60mm at the fingers. The hand is only 25mm thick. The hands are tendon driven, with most of the motors located in the forearm. Simulations indicated that for crawling, sitting and squatting a 5 DOF leg is adequate. However, it was decided to incorporate an additional DOF at the ankle to support standing and walking. Therefore each leg has 6 DOF: 3 DOF at the hip, 1 DOF at the knee, and 2 DOF at the ankle (flexion/extension and abduction/adduction). The foot twist rotation is not implemented. Crawling simulation analysis also showed that a 2 DOF waist/torso is adequate. However, to support manipulation, a 3 DOF waist has been incorporated.

[4] Current information on the iCub project is available at www.icub.org.

Fig. 7.2 The iCub hand has three independent fingers (including the thumb) while the fourth and fifth fingers are used for additional stability and support and have one DOF overall

This provides increased range and flexibility of motion for the upper body resulting in a larger workspace for manipulation (e.g. when sitting). The neck has a total of 3 DOF and provides full head movement. The eyes have further 3 DOF to support both tracking and vergence behaviours.

From the sensory point of view, the iCub is equipped with digital cameras, gyroscopes and accelerometers, microphones, and force/torque sensors. A distributed sensorized skin is under development using capacitive sensor technology. Each joint is instrumented with positional sensors, in most cases using absolute position encoders. A set of DSP-based control cards, custom-designed to fit the iCub, takes care of the low-level control loop in real-time. The DSPs communicate with each other via a CAN bus. Four CAN bus lines connect the various segments of the robot. All sensory and motor-state information is transferred to an embedded Pentium-based PC104 computer which acts as a hub and handles synchronization and reformatting of the various data streams. This hub then communicates over a Gbit ethernet cable with an off-board computer system which takes responsibility for the iCub's high-level behavioural control (see Fig. 7.3). Because the iCub has so many joints to be configured and such a wealth of sensor data to be processed, the iCub software has be configured to run in parallel on a distributed system of computers to achieve real-time control. This in turn creates a need for a suite of interface and communications libraries — the iCub middleware — that runs on this distributed system, effectively hiding the device-specific details of motor controllers and sensors and facilitating inter-process and inter-processor communication. We discuss the iCub middleware in the next section.

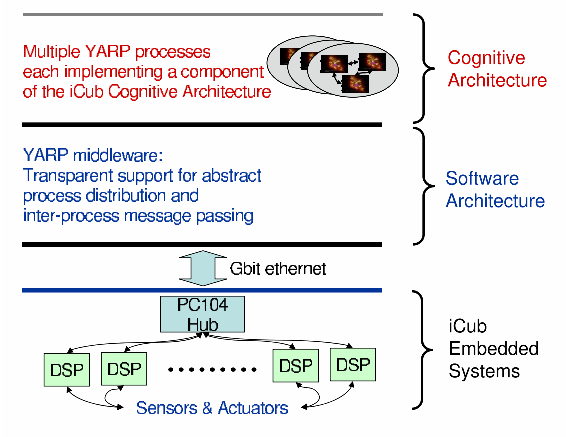


Fig. 7.3 The layers of the iCub architecture: the cognitive architecture and software architecture run on multiple computers located remotely from the iCub robot. The PC104 hub and the DSPs are located on the iCub. Communication between these two sub-systems is effected by a 1Gbit ethernet connection. The power supply is also located remotely from the iCub. Both power and communications are delivered by an umbilical cord.

7.1.2 The iCub Middleware

The iCub software is developed on top of YARP [96], a set of libraries that supports modularity by abstracting two common difficulties in robotics: algorithmic modularity and hardware interfacing. The YARP libraries assume that a real-time layer is in charge of the low-level control of the robot and instead they take care of defining a soft real-time communication layer and hardware interface that is suited for cluster computation. YARP also takes care of providing independence from the operating system and the development environment. The main tools in this respect are ACE [172] and CMake. The former is an OS-independent communication library that hides the problems of interprocess communication across different operating systems. CMake is a cross-platform make-like description language and tool to generate appropriate platform specific project files.

YARP abstractions are defined in terms of protocols. The main YARP protocol addresses inter-process communication issues. The abstraction is implemented by the port C++ class. Ports follow the observer pattern by decoupling producers and consumers. They can deliver messages of any size across a network using a number of underlying protocols (including shared memory when possible). In doing so, ports decouple as much as possible (as function of a certain number of user-defined parameters) the behavior of the two sides of the communication channels. Ports can be commanded at run time to connect and disconnect.

The second abstraction of YARP concerns hardware devices. The YARP approach is to define interfaces for classes of devices which wrap native code APIs (often provided by the hardware manufacturers). Changes in hardware will most likely require only a change in the API calls (and linking against the appropriate library). This easily encapsulates hardware dependencies but leaves dependencies in the source code. The latter can be removed by providing a "factory" for creating objects at run time (on demand). The combination of the port and device abstractions leads to remotable device drivers which can be accessed across a network: e.g. a grabber can send images to a multitude of listeners for parallel processing.

Overall, YARP's philosophy is to be lightweight and to facilitate the use of existing approaches and libraries. This naturally excludes hard real-time issues that have to be necessarily addressed elsewhere, typically at the operating system level.

Higher-level behaviour-oriented application sofware typically comprises several coarse-grained YARP processes. This means that to run iCub applications, you simply need to invoke each process and instantiate the communication between them. The YARP philosophy is to decouple the process functionality from the specification of the inter-process connections. This encourages modular software with reusable processes that can be used in a variety of configurations which are not dependent on the functionality of the process or embedded code.

7.2 The iCub Cognitive Architecture

The design and realization of any cognitive architecture is a long-term project. The iCub cognitive architecture is no different. Inevitably, the realization happens in stages. In what follows, we describe a cognitive architecture which follows a significant subset of the roadmap guidelines to a greater or lesser extent. The goal of this preliminary cognitive architecture is to integrate some of the phylogenetic capabilities identified in Chap. 6 in a way that is meaningful for both neurophysiology and developmental psychology. In other words, the current goal is to build a minimal but faithful functioning system as a proof of principle. Sect. 7.3 discusses the degree to which each guideline has been followed and Chap. 8 considers the challenges posed by a complete implementation of the roadmap guidelines.

On the basis of the main conclusions drawn in Sect. 6.1.5, we decided to focus on several key capabilities. Gaze control, reaching, and locomotion constitute the initial simple goal-directed actions. Episodic and procedural memories are included to effect a simplified version of internal simulation in order to provide capabilities for prediction and reconstruction, as well as generative model construction bootstrapped by learned affordances.

In addition, motivations encapsulated in the system's affective state are made explicit so that they address curiosity and experimentation, both explorative motives, triggered by exogenous and endogenous factors, respectively. This distinction between the exogenous and the endogenous is reflected by the need to include an attention system to incorporate both factors.

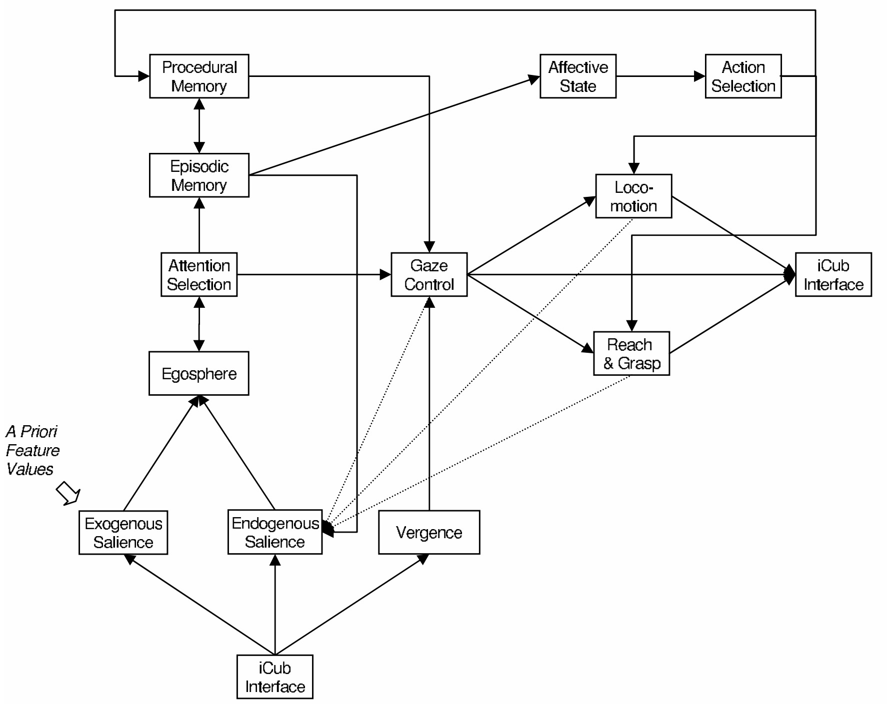


Fig. 7.4 The iCub cognitive architecture

A simple process of homeostatic self-regulation governed by the affective state provides elementary action selection. Finally, all the various components of the cognitive architecture operate concurrently so that a sequence of states representing cognitive behaviour emerges from the interaction of many separate parallel processes rather than being dictated by some state-machine as in the case of most cognitive architectures. This preliminary cognitive architecture comprises thirteen components, as follows (refer also to Fig. 7.4).

1. Exogenous Salience
2. Endogenous Salience
3. Egosphere
4. Attention Selection
5. Episodic Memory
6. Procedural Memory
7. Affective State
8. Action Selection
9. Gaze Control

10. Vergence
11. Reach & Grasp
12. Locomotion
13. iCub Interface

Together, the Exogenous Salience, Endogenous Salience, Egosphere, and Attention Selection components comprise the iCub's perception system. Similarly, Gaze Control, Vergence, Reach & Grasp, Locomotion comprise the iCub's actions system. The Episodic Memory and the Procedural Memory together provide the iCub's principle mechanism for anticipation and adaptation. The Affective State component effects the iCub motivations which together with the Action Selection component provide a very simple homeostatic process which regulates the autonomous behaviour of the iCub. The iCub Interface component completes the architecture and reflect the embodiment of the iCub from an architecture point of view. Thus, we can map each cognitive architecture component to one of the seven headings under which we grouped the roadmap guidelines in the previous chapter. For that reason, we will structure the discussion of each component in the same way, under these seven headings, both in terms of the description of the operation of each component and, at the end of the chapter, in terms of the extent to which they fulfil the requirements of the roadmap guidelines.

We now address the specification of each component and the implementation of the cognitive architecture as an interconnected set of YARP software modules.[5]

7.2.1 Embodiment: The iCub Interface

The iCub robot is realized as an embodied agent by its physical components: links, joints, actuators, and a variety of sensors providing proprioceptive and exteroceptive information about the iCub body and its local environment. Actuation is effected through a variety of brushless motors, DC motors, and servo-motors. Sensory data includes joint position, velocity, and torque, streamed video images from the left and right eye cameras, audio from the left and right ear microphones, and 3-D head attitude and acceleration from the inertial sensor. In addition, tactile images from the iCub's finger-tips, arms, and body are also available.

Access to the motor control and sensor interface is provided through a standard iCub component denoted the iCub Interface in the cognitive architecture (see Fig. 7.4). This module is instantiated as an YARP software module `iCubInterface2`. In general, it provides an abstract view of the iCub's motors and sensors, facilitating control and data acquision through YARP ports. The names of the ports instantiated by iCub Interface adhere to a standard format that names a specific iCub robot and divides the robot into six parts. These parts are the `head`, `torso`, `right_leg`, `left_leg`, `right_arm`, and `left_arm`. For each part, there are three ports:

[5] Throughout the following, when we speak of some element of the cognitive architecture, we will refer to it as a component of the architecture and when we speak of its software instantiation, we will refer to it as a YARP software module (or set of modules).

1. `/robotname/part/rpc:i`
2. `/robotname/part/command:i`
3. `/robotname/part/state:o`

where `robotname` is the name of the robot (e.g. `icub`), and `part` is the logical group of joints referred to by the port. The `rpc:i` port provides a textual command interface to all of the devices defined in the part. The `command:i` port receives a stream of vectors which contains velocity commands. In general this is useful for closed loop, servoing control. The `state:o` port streams out a vector containing the motor encoder positions for all joints.

Table 7.1 Mapping between joint and joint numbers

Body Part	**Joint Description**	**Joint Number**
Head	Neck pitch	0
	Neck roll	1
	Neck yaw	2
	Eyes tilt	3
	Eyes version	4
	Eyes vergence	5
Left/Right Arm	Shoulder pitch	0
	Shoulder roll	1
	Shoulder yaw	2
	Elbow	3
	Wrist prosupination	4
	Wrist pitch	5
	Wrist yaw	6
	Hand finger adduction/abduction	7
	Thumb opposition	8
	Thumb proximal flexion/extension	9
	Thumb distal flexion	10
	Index proximal flexion/extension	11
	Index distal flexion	12
	Middle proximal flexion/extension	13
	Middle distal flexion	14
	Ring and little finger flexion	15
Torso	Torso yaw	0
	Torso roll	1
	Torso pitch	2
Left/Right Leg	Hip pitch	0
	Hip roll	1
	Hip yaw	2
	Knee	3
	Ankle pitch	4
	Ankle roll	5

In each iCub part, the joints are numbered to give a natural open kinematic chain, with the base reference frame on the torso. The most proximal joint is denoted joint 0 and the most distal joint is denoted N_{max}. The key reference point on the body is the base of the neck. The joint numbers are used when controlling joint motors or acquiring joint sensor data. For example, the `head` part has 6 joints which determine the head attitude and eye gaze. Joint numbers 0, 1, 2, 3, 4, and 5 correspond to the neck pitch, neck roll, neck yaw, common eye tilt, common eye version, and common eye vergence, respectively. The mapping between joint numbers and joint names is summarized in Table 7.1. Sending the following text string `set pos 0 45` to the `/icub/head/rpc:i` port will command axis 0 of the head (neck pitch) to 45 degrees. Furthermore, the iCub's facial expressions, such as eye-lid position, eyebrow shape, and mouth shape can be controlled using a different port and control protocol.

In addition to the motor encoder position, other sensory data is also available on specific iCub YARP ports. These include the two cameras comprising the iCub's eyes, the two microphones comprising the iCub's ears, the inertial sensor comprising the iCub's vestibular system, one six-axis force/torque sensor on each of the two

Table 7.2 Standard port names for iCub actuators and sensors

Actuator	Port Name	Comment
Joint motors	/icub/head/command:i /icub/left_arm/command:i /icub/right_arm/command:i /icub/torso/command:i /icub/left_leg/command:i /icub/right_leg/command:i	Streamed vector of motor velocities in degrees/s
Facial expression	/icub/face/raw /icub/face/emotions	low level interface high level interface
Sensor	**Port Name**	**Comment**
Motor encoders	/icub/head/state:o /icub/left_arm/state:o /icub/right_arm/state:o /icub/torso/state:o /icub/left_leg/state:o /icub/right_leg/state:o	Streamed vector of motor positions in degrees
Cameras	/icub/cam/left /icub/cam/right	Streamed RGB images
Microphones	/icub/mics	
Inertial sensor	/icub/inertial	Streamed vector of 3-axis orientation, angular velocity, acceleration values
Force/torque	/icub/left_arm/analog:o /icub/right_arm/analog:o	Streamed vector of 3 force and 3 torque values
Skin	/icub/skin/lefthand /icub/skin/righthand	Streamed vector of 102 tactile sensor readings in each hand
Finger encoders	/icub/left_hand/analog:o /icub/right_hand/analog:o	Streamed vector of 15 fi nger joint positions in degrees

arms, tactile sensors on the iCub's finger-tips, and position sensors for all 15 finger joints in each hand. The YARP port names for all control and sensor data acquisition are summarized in Table 7.2.

7.2.2 Perception

We noted already that the iCub provides both proprioceptive and exteroceptive sensory data and we explained in the previous section how this data is accessed through the iCub Interface component using YARP ports. In this section, we focus on the components of the cognitive architecture that deal with the processing and interpretation of exteroceptive sensing, in general, and of visual and aural sense data, in particular. The core of the iCub exteroceptive perceptual capabilities is a multimodal saliency-based attention system described in [328] and from which much of this section was abstracted.

The visual part of this attention system is based on the Itti and Koch model of selective visual attention [178, 180, 199] while the aural part effects binaural sound localization based on both temporal and spectral analysis [169]. The visual and acoustic salience maps are combined in a continuous spherical surface in a mapping from 3D point locations to an ego-sphere representation centred at the iCub's body.

This attention system uses local conspicuity feature values to determine the salience of regions in the image and it adopts a winner-take-all strategy to determine the most salient region from among all candidates in the salience map. The position of the winner determines the shift in gaze so that the robot eyes are subsequently directed towards this most salient region.

Salience-based attention is often a completely context-free process, depending only on the feature values present in the sensory stimulus to determine salience. We refer to this as exogenous salience.

However, some models can be adapted to allow the inclusion of context-aware sensitivity of the attention system. Typically, they do this in two ways. The first way is to allow the feature values that are considered to be salient to be specified or weighted either by fixed weights or by weights determined during a training phase. The second way is by explicit inclusion of an endogenous salience map which is modulated directly by the similarity of image regions to some a priori target landmark feature values or appearance. This endogenous salience is determined elsewhere in the system. Both of these strategies are adopted in the iCub architecture. The feature value tuning is effected through a user-guided interactive selection of the feature values that are associated with target and landmark objects. The endogenous salience is effected by the system's episodic memory which exploits a scale-invariant representation of the appearance of landmarks it has encountered as it explored its environment. At present, the iCub cognitive architecture does not exploit endogenous aural salience.

Once the visual and aural salience maps have been combined in the egosphere, the final salience map is subjected to a winner-take-all non-maxima suppression process to identify the most salient location. A second inhibition-of-return mechanism modulates the ego-sphere salience map to generate a temporal scan-path which ensures that the iCub's attention is not always fixed on the same location and is predisposed to exploring its local environment. Both of these processes occur in the Attention Selection component. This component then feeds the relevant coordinates to the Gaze Control component which then directs the iCub's gaze accordingly so that it fixates on some point of interest in its environment. A separate process in the *Vergence* component adjusts the gaze angles of the left and right camera so that their principal rays, i.e. the line through the image centre and the focal point, intersect at this point.

7.2.2.1 Exogenous Salience

We distinguish between selective attention that is driven by external stimuli and attention that is driven by the state of the system itself. The latter is often referred to as top-down visual attention to reflect a processing model comprising low-level feature extraction and high-level feature interpretation. High-level processes typically assume the use of more explicit, and often symbolic, information. The salience of a top-down attentional system then is often guided by some knowledge-based understanding of the scene. Our viewpoint on attention, in particular, and system processing, in general, takes a different stance. Rather than adopting a high-level vs. low-level dichotomy, we take the view that system organization is much flatter with no semantic hierarchy but with several reentrant loops. Thus, salience can be influenced not just by high-level knowledge but by any aspect of the system state, such as episodic memory and the current state of the system's actuators (and actions). For example, attention might be conditioned by gaze or by locomotion, as much as by memory. We adopt the term 'endogenous salience' to reflect this dependence on internal state (in contradistinction to top-down knowledge). Conversely, the term 'exogenous salience' reflects a dependence on the content of the stimuli themselves, rather than the state of the system and, in particular, exogenous visual salience is determined by the contrast of visual features of the image.

As noted already, for the iCub cognitive architecture, we have adopted and adapted the salience-based model of selective visual attention proposed by Itti and Koch [177, 178, 180]. It comprises a pre-attentive phase and an attentive phase. The pre-attentive phase is concerned principally with the processing visual stimuli to extract a variety of visual features and their subsequent combination into a saliency map. The attentive phase then uses this information to determine the location to which the attention should be drawn [107, 179, 199]. It typically exploits two control strategies: a Winner Take All (WTA) process to identify a single focus of attention, and an Inhibition Of Return (IOR) mechanism to diminish the attention value of a winning location so that other regions become the focus of attention. The general approach adopted by Itti and Koch [177, 178, 180], based on earlier work by Koch and Ullman [199], is to process an colour image of a scene, extract several features (such as intensity, double-opponent colour, and local orientation) at several scales, perform centre-surround filtering to establish local contrast, and then normalize the resulting feature maps. The intensity features are extracted with

Mexican hat wavelets of different sizes. In our version for the iCub, the colour saliency feature follows the implementation in [47]. Directional saliency is computed using Gabor filters at three different scales $\sigma \in 2.0, 4.0, 8.0$ in four different directions $\theta \in 0°, 45°, 90°, 135°$. Every scale of each feature map is then recombined and re-normalized to generate a so-called conspicuity map for each feature. These conspicuity maps are then combined linearly to create a single saliency map. Following Koch and Ullman's note that other features may also be involved in the generation of the saliency map [199, 179], we have incorporated an additional feature channel in the form of visual motion detected using the Reichardt correlation model [304]. Stereo disparity, and vergence-driven zero stereo disparity in particular, is being considered as a future feature channel.

Auditory salience requires the localization of a sound source in the iCub's acoustic environment. The position of a sound source is estimated using interaural spectral differences (ISD) and interaural time difference (ITD) as described in detail in [169]. The localization process makes use of two specially-formed ears, each having a spiral pinna. This form of artificial pinna gives spectral notches at different frequencies depending on the elevation of the sound source. A notch is created when the sound wave reaching the microphone is directly canceled by the wave reflected in the pinna. This happens when the microphone-pinna distance is equal to a quarter of the wavelength of the sound (plus any multiple of half the wavelength). The notches, and with it the elevation estimate, can then be found by looking for minima in the frequency spectra. To obtain the azimuth angle of the sound source, we use the interaural time difference. By looking for the maximum cross correlation between the signals at the right and left ear it is easy to calculate the ITD and the azimuth angle knowing the distance between the microphones and the sampling frequency. To get a measurement of the uncertainty in the estimated position we divide the samples in several subsets and calculate the azimuth and elevation angle for each subset. We then compute the mean and standard deviation of the estimates.

7.2.2.2 Endogenous Salience

In the current version of the iCub cognitive architecture, endogenous salience is determined in a simple manner based on the contents of episodic memory (see Sect. 7.2.4.1) using colour segmentation. Specifically, the current image acquired by the left eye of the iCub (assumed to be the dominant eye) is segmented into foreground and background, the largest blob in the foreground is identified, and the centroid of this blob is used to define the location of the most salient region. The colour segmentation is carried out in HS (Hue-Saturation) space. The parameters for the segmentation, i.e. the hue and saturation ranges of foreground regions, are extracted automatically from a log-polar image which is provided as input to the Endogenous Salience component by the Episodic Memory. Typically, this log-polar image will have been generated by the attention system when fixating on some point of interest when exploring or interacting with its environment. The hue and saturation ranges are defined as the log-polar image modal hue and saturation value plus or minus some tolerance (expressed as a percentage of the hue and saturation range values).

Since we use log-polar images, the hue and saturation values are heavily biased to the scene content at the centre of the image and, thus, to the focus of attention when the image was acquired and stored in the episodic memory. The segmented image is filtered to remove small regions by performing a morphological opening. The endogenous salience component communicates the location of the most salient point directly to the egosphere component.

7.2.2.3 Egosphere

The ego-sphere is a projection surface for spatially-related information which is used in the iCub cognitive architecture to build a coherent representation of multi-modal saliency. The ego-sphere is head-centered and fixed with respect to the robot's torso. In this way, rotational or translational corrections are not required as long as the robot does not translate. For efficiency reasons, the egocentric maps are stored as rectangular images, expressed in spherical coordinates (azimuth ϑ and elevation φ). The map's origin ($\vartheta = 0°$, $\varphi = 0°$) is located at the center of the image. Saliency information is projected onto the egosphere by first converting the stimulus orientation to torso-based, head-centered frame of reference in spherical coordinates using head kinematics and then projecting the stimulus intensity onto a modality-specific rectangular egocentric map (see in [328] for a more detailed explanation).

After converting the saliency information to a common egocentric frame of reference, we need to combine the different sensory modalities into a single final map. This is done by taking the maximum value across all saliency channels at each location. This simple approach is fast and works well in many situations. However, the cognitive architecture would benefit from more sophisticated context-sensitive multi-modal approaches, especially ones which learn the saliency aggregation function from experience.

The egocentric saliency map also acts as a form of short-term memory for salient regions. This allows the iCub to behave in a more sophisticated way than a purely reactive agent would. By using memorized saliency information, the iCub can shift attention to previously-discovered salient regions. This is achieved by allowing maintained salience values which have been mapped to the egosphere allowing them to decay with time.

7.2.2.4 Attention Selection

Based on the saliency ego-sphere, the iCub is able to select points of interest in order to explore its surroundings by using basic attentional mechanisms, cued either by exogenous or endogenous stimuli. This exploratory behaviour results from the interplay between the attention selection and inhibition mechanisms. The attention selection process selects a point corresponding to the location with the highest overall saliency value. The inhibition mechanism attenuates the saliency for locations which have been close to the focus of attention for a certain amount of time. This modulation of the saliency map is motivated by the inhibition of return (IOR) observed in human visual psychophysics [297]. To realize inhibition of return, two additional egocentric maps are used, the habituation map H and the inhibition map A. The habituation map encodes the way the visual system gets used to persistent or repetitive stimuli, with the system's attention being drawn toward more novel salient regions in the environment. In our system, the habituation map is initialized to zero and updated according to a Gaussian weighting function that favours the regions closer to the focus of attention. While attending to a salient point, the habituation map at that location will asymptotically tend to 1, with a parameter d_h determining the speed of convergence.

Whenever the habituation value exceeds a predefined level, the system becomes attracted to novel salient points. The inhibition map represents perceptual regions that have already been attended to and that should not capture the system's attention again in the near future, i.e., local habituation leads to inhibition of that location [297].

The inhibition map is initialized to 1.0. When the habituation of a certain region exceeds a threshold $t_h = 0.85$, the inhibition map is modified by adding a scaled Gaussian function, G_a, with amplitude -1, centered at region with maximum habituation. The resulting effect is that a smooth cavity is added to the inhibition map at the relevant position. The temporary nature of the inhibition effect is modeled by a time-decay factor, $d_a \in [0, 1]$, applied at every location of the inhibition map.

Finally, for attention selection the multimodal saliency map is multiplied by the inhibition map, thus combining instantaneous saliency and the memory of recently attended locations.

By combining saliency based attention selection and the inhibition mechanism, saccadic eye movements toward the most salient locations occur in a self-controlled way. The inhibition mechanism ultimately causes the formation of a stimulus-driven exploratory behavior. The ability to attend to specific region of interest for a longer time period can be achieved by adapting the habituation gain/threshold when attention is based on endogenous salience, although this has not yet been implemented in the iCub cognitive architecture.

7.2.3 Action

7.2.3.1 Gaze Control

The Gaze Control component provides coordinated control of the head and eyes of the iCub. Rather than by specifying the raw joint values for the head and eyes, the gaze direction of the robot is specified by gaze direction (azimuth and elevation angles) and vergence of the two eyes. In addition, the motion of the eyes and head when moving to a given gaze position are controlled to provide a motion profile that is similar to humans, with the gaze direction of the eyes moving quickly and the head moving more slowly but subsequently catching up so that the eyes gaze is reaching an equilibrium position that is relatively centred with respect to the head (allowing for the specified vergence) and with the head oriented in the specified gaze direction.

The gaze direction can be specified using any of four different types of coordinates:

1. the absolute azimuth and elevation angles (in degrees);
2. the azimuth and elevation angles relative to the current gaze direction;
3. the normalized image coordinates (in the range zero to one) of the position at which the iCub should look;
4. the image pixel coordinates of the position at which the iCub should look.

Head gaze and vergence are controlled simultaneously.

There are two modes of head gaze control: saccadic motion and smooth pursuit motion. Saccades are controlled in two phases. In the first fast phase, the eyes are driven quickly to the destination gaze direction by controlling the common eye version and tilt degrees of freedom. In the second slow phase, the neck moves toward the final gaze direction with a slower velocity profile and the eyes counter-rotate to keep the image stable. A new saccade is accepted only when the previous one has finished. This is the typical operation in humans where saccades are used to change the object of interest [226]. Smooth pursuit only operates in the slow phase, but it accepts a continuous stream of commands. It is meant to emulate the human behaviour when tracking an object. A single set of motor controller gains are used for both the saccade and smooth pursuit modes. These gains specify the speed of the motion and therefore the amount of motion undergone by the eyes and by the neck. The higher the gain, the more the eye motion will lead the neck motion and, therefore, the more the eyes have to counter-rotate as the neck approaches the gaze direction. Vergence operates continuously in a single phase and there is an independent gain for the vergence controller.

7.2.3.2 Vergence

Vergence control refers to the adjustment of the relative orientation of a pair of cameras so that the optical axis of each camera intersects at a given point of interest in the scene. This in turn implies that the scene content in the region around the optical axis is more or less identical. There will inevitably be some slight differences in appearance due to the fact that the region is being viewed from two different positions. The smaller the base-line separating the two cameras, the smaller these differences will be. Thus, one can view vergence control as an exercise in active local image registration with the goal of keeping the central areas of the two images aligned.

Registration techniques typically use one of several classes of algorithm based on image correlation or feature matching in the space domain and Fourier methods in the spatial-frequency domain [301]. If the baseline between the two cameras is sufficiently small, as is the case in the iCub robot where it is the same as the interocular distance of an human child, then the transformation required to register regions in the image can be approximated by a simple translation instead of a more complex affine transformation or non-linear warping. In the case of a binocular stereo camera configuration where there is only one degree of freedom in the relative transformation between the cameras, it is sufficient to measure just the horizontal translation required to register the left and right images.

The translation necessary to effect registration can be determined in several ways. One popular approach is to construct the stereo disparity map. However, this can be computationally expensive to compute and requires subsequent analysis of the resultant disparity image to establish the distinct regions of similar disparity. We would prefer to compute the translation, and hence the camera rotation, directly. This is accomplished using the Fourier cross-power spectrum. Although the cross-power spectrum has been used in the past to identify the unique translation required to register two images, it is often the case that there is more than one point of interest in the local region around the optical centres of the camera and these points may be at different distances from the two cameras. Consequently, there may be more than one possible translation in a given local region. The iCub Vergence component uses the Fourier cross-power spectrum of suitably apodized images[6] to identify the multiple translations required to register different parts of a local region and thereby control camera vergence. This requires some criterion for selecting one of candidate translations to control the vergence. In the current set-up, the translation corresponding to the foreground region, i.e. the point of interest closest to the two cameras, is selected.

The Vergence component outputs the disparity of the selected region, which is effectively the position of the corresponding maximum, in normalized coordinates (-1, +1). This output is communicated to the Control Gaze component which effects the actual control function.

7.2.3.3 Reach and Grasp

The Reach & Grasp component includes a robust task-space reaching controller for learning internal generic inverse kinematic models and human-like trajectory generation. This controller takes into account various constraints such as joint limits, obstacles, redundancy and singularities. It also includes a module for grasping based on reaching and orienting behaviours. This allows the coordination of looking (for a potential target), reaching for it (placing the hand close to the target) and attempting a grasping motion (or another basic action).

At present, the reaching controller takes as input the Cartesian position and orientation of an object to be grasped, based on a visual recognition process. The arm and torso configuration to achieve the desired pose (i.e. the hand position and orientation) is determined using a non-linear optimizer which takes into account all the various constraints, one of which is to keep the torso as close as possible to vertical while reaching. Subsequently, human-like quasi-straight hand/arm trajectories are then generated independently using a biologically-inspired controller. This controller exploits a multi-referential system whereby two minimum-jerk velocity

[6] A Gaussian function is used to apodize or window the images and thereby define the extent of the local region of interest around the cameras' optical centres which is considered in the Fourier cross-power spectrum.

vectors are generated, one in joint space and one in task space [142]. A coherence constraint is imposed which modulates the relative influence of the joint and task space trajectories. The advantage of such a redundant representation of the movement is that a quasi-straight line trajectory profile which is similar to human-like motion can be generated for the hand in the task space while at the same time retaining convergence and robustness against singularities.

In the future, we plan on eliminating the Cartesian representation from this approach and using the gaze parameters directly. Since the state of the motors that control the motion of the robot are dependent on the state of the motors that control the head and gaze, this form of robot control has been dubbed motor-motor control to distinguish it from sensory-motor control [250].

The act of grasping an object requires one to match the hand position, orientation, and shape to that of the object, in a way that reflects the intended use of the object. Thus, grasping involves a relationship between perceived and perceiver in the context of action. Grasping is therefore intimately related to the concept of affordance, J. J. Gibson's term for the action possibilities of an object with reference to the actor's capabilities. Affordances reflect the goals, interests, and capabilities of the perceiver as much as it reflects the characteristics of the object itself. As we saw in Chap. 4, this applies equally to grasping, with grasping actions being goal-directed and selected in part by the appearance (size, shape, and orientation) of the object to be grasped.

The iCub learns visual descriptions of grasping points from experience [255]. The essence of the approach is to compute a map between visual features and the probability of good grasp points. The information can be used to look actively for good grasp points in new object by reusing experience with previous objects and thus direct exploration.

We return to the issue of affordances in Chap. 8 where we discuss future work which will focus on the integration of the iCub's stand-alone ability to learn affordances [256, 257, 258] with the procedural memory component of iCub cognitive architecture to be described below.

7.2.3.4 Locomotion

The purpose of the Locomotion component is to effect the movement of the iCub towards some goal location. We distinguish here between navigation and locomotion. Navigation is concerned with the strategy of how to get from one point in the environment to another, planning an effective and efficient path which avoids obstacles, which doesn't lead to dead-ends, and which minimizes the cost of the total travel, measured in, for example, time, energy, or distance. Locomotion, on the other hand, simply specifies how to control the robot motors to move the robot towards some well-specified target location. As a process, locomotion is local in time and space — i.e. it considers only relatively small motions – whereas navigation is extended in time and space.

The iCub was originally designed to effect locomotion by crawling on hands and knees, just like a young child, although sufficient degrees of freedom have been incorporated to allow bi-pedal walking. Locomotion by crawling, then, requires a controller that uses central pattern generators that produce motor primitives implemented as coupled dynamical systems to produce stable rythmic and discrete trajectories [79, 306, 307, 308]. This controller includes four different simple behaviours:

1. crawling;
2. transition from sitting to crawling;
3. transition from crawling to sitting;
4. reaching for a target object.

These different actions are used to implement a more complex behaviour where the iCub moves towards a target object identified by vision. This is accomplished using a planner based on force fields which steers the iCub towards the target while avoiding obstacles in its way. When the iCub is close enough to a target object, it stops and reaches for it. While the transition from crawling to sitting (and vice versa) has been successfully implemented in simulations, some current limitations on the iCub joints have temporarily prevented it being realized on the physical robot.

7.2.4 Anticipation & Adaptation

We have emphasized throughout this book 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 Sect. 7.3.2 our sensory apparatus should 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 to 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' [37].

This internal simulation in important for accelerating the scaffolding of 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.

Internal simulation works concurrently with the other component capabilities and, in fact, the simulation circuitry provides just another input to the action selection process which modulates the dynamics of the architecture. 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'.

A 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 [400]: the simulation can be effected faster than real-time. The iCub cognitive architecture uses a combination of episodic memory and procedural memory to accomplish both anticipation and adaptation.

7.2.4.1 Episodic Memory

The Episodic Memory component is a simple memory of autobiographical events. It is a form on one-shot learning and, in its present guise, does not generalize multiple instances of an observed event. That functionality needs to 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. The current version implements a simple form of content-addressable memory based on colour histograms [365, 366] and log-polar mapping [38, 39, 44]. 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 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 log-polar images as the invariant landmark representation [38, 39, 44] and matching is effected using colour histograms and (a variant of) colour histogram intersection [365, 366]. 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, the stored image is recalled; otherwise, the presented image is stored. In principle, the images presented to the module can be either conventional Cartesian images or log-polar mapped images. However, we use log-polar images in the iCub cognitive architecture for their scale and rotation invariance properties and because they are effectively centre-weighted due to the non-linear sampling and low-pass filtering at the periphery. This makes it possible to effect appearance-based image/object recognition without prior segmentation.

7.2.4.2 Procedural Memory

The Procedural Memory component is a network of associations between action events and pairs of perception events. In the current version of the iCub cognitive architecture, 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. As noted above, this generalization needs to be effected at some future point by a 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.

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. In the first mode, the image perception presented to the memory represents the first perception P_i 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. In the second case, the image perception presented to the memory represents the second perception P_k 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 $(P_i, \sim, \sim)$ or $(\sim, \sim, P_k)$.

The use of episodic and procedural memories to realize internal simulation differs from other approaches. For example, Shanahan uses paired hetero-associative memories which encapsulate sensory and motoric representations in two distinct components, with the relationship between them being captured implicitly in the dynamics of their mutual interaction [337, 338, 336, 335] (see also Appendix A, Sect. A.2.2). On the other hand, the episodic memory encapsulates the sensory representations while the procedural memory makes explicit the relationship between the sensory stimuli and associated actions.

7.2.5 Motivation: Affective State

The Affective State component is a competitive network of three motives:

1. curiosity (dominated by exogenous factors);
2. experimentation (dominated by endogenous factors);
3. social engagement (where exogenous and endogenous factors balance).

The outcome of this competition fed to the Action Selection component which produces a cognitive behaviour that is biased towards learning (motivated by curiousity), and prediction and reconstruction (motivated by curiousity and social engagement). Both curiousity and experimentation are forms of exploration. Social engagement is an exploratory act to but it focusses on the establishment of a common basis for mutual interations with others rather than on acquiring new knowledge about the world. In a sense, exploration concerns the acquisition of knowledge, whereas social engagement concerns the agreement of knowledge.

In the current version of the iCub cognitive architecture, we implement only simple forms of curiousity and experimentation motives. Both motives are modelled as a temporal series of event-related spikes. The current level of a motivation is a weighted sum of recent spiking activity, with weighting being biased towards most recent spikes. Specifically, the weighting function decreases linearly from 1 to 0 as a function of the time. The number of time intervals over which the sum is taken is provided as a parameter to the software module which realizes this component.

Curiousity and experimentation have different spiking functions. A curiousity spike occurs when either a new event is recorded in the episodic memory or when an event that has not recently been recalled is accessed in the episodic memory. The number of time slots that determine the expiry of an event (and hence whether or not is has been recently recalled) is provided as a parameter. An experimentation spike occurs when an event / image which is predicted or reconstructed by the procedural memory and recalled by the episodic memory is subsequently recalled again by the episodic memory, typically as a result of the endogenous salience successfully causing the attention system to attend to the predicted / reconstructed event. This is equivalent to the sequential recall of the same event in episodic memory when in either prediction or reconstruction action mode.

The Affective State component receives inputs from Episodic Memory component and the Action Selection component. The input from Episodic Memory includes the identification of the previously accessed and currently accessed events / images in the memory. The input from Action Selection includes the current curiousity level, current experimentation level, and their respective instantaneous rates of change.

7.2.6 Autonomy: Action Selection

The purpose of the Action Selection component is to effect the development of the iCub and specifically to increase its predictive capability. In the current version of the iCub cognitive architecture, this means the selection of the iCub's mode of exploration. At present, only two basic motives drive this iCub development, curiousity and experimentation, both of them exploratory. The third main motive, social interaction, has not yet been addressed. Consequently, the present implementation of action selection is based on a very trivial function of the levels of curiousity and experimentation produced by the Affective State component. Specifically, the learning mode is selected if the curiosity level is higher than the experimentation level; otherwise the prediction mode is selected.

7.2.7 Software Implementation

The architecture as shown in Fig. 7.4 has been realized as a complete software system comprising a suite of YARP iCub modules. These modules comprise the following.

- *salience*
- *endogenousSalience*
- *egoSphere*
- *attentionSelection*
- *episodicMemory*
- *proceduralMemory*
- *affectiveState*
- *actionSelection*
- *controlGaze2*
- *crossPowerSpectrumVergence*
- *logPolarTransform*
- *cameraCalib*

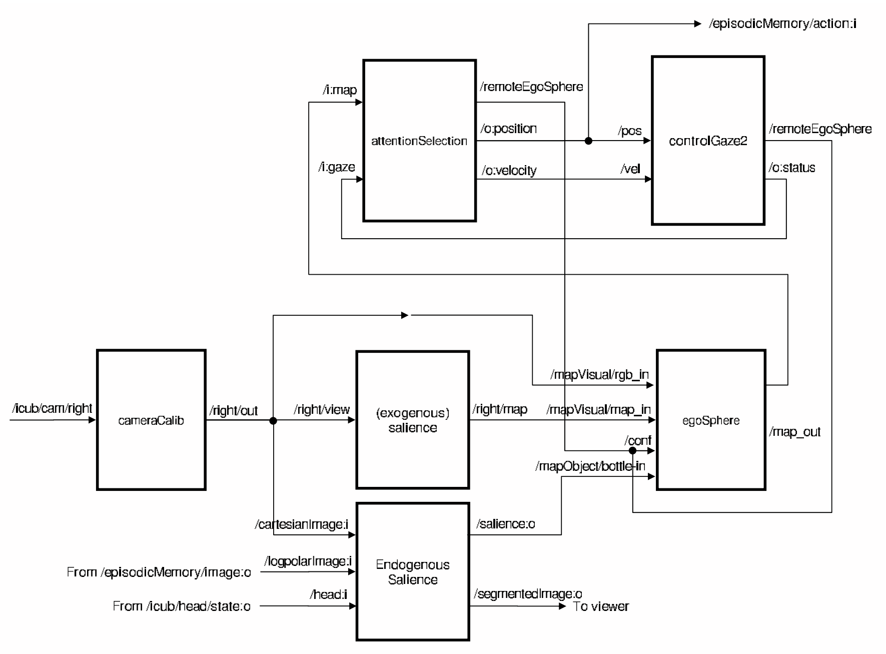


Fig. 7.5 Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part A)

These modules are integrated as a single iCub application which is shown schematically in Figures 7.5 and 7.6. There is a clear one-to-one correspondence between the iCub cognitive architecture components in Fig. 7.4 and the YARP modules in Figures 7.5 and 7.6. Note that Reach & Grasp and Locomotion architecture components have not yet been integrated in the suite of YARP modules although they have been implemented on a stand-alone basis. The iCub Interface component is simply a catch-all place-holder for the YARP interfaces to the various motors and sensors and so does not appear in Figures 7.5 and 7.6. Two additional modules appear in Figures 7.5 and 7.6 which don't feature in Fig. 7.4. These are the *logPolarTransform* and *cameraCalib*, utility modules that handle camera calibration and image transformation between Cartesian and log-polar formats, respectively. The interconnections between the YARP modules are the port connections by which the YARP modules communicate with each other. Each interconnection is named as either an input port to or an output port from a given module.

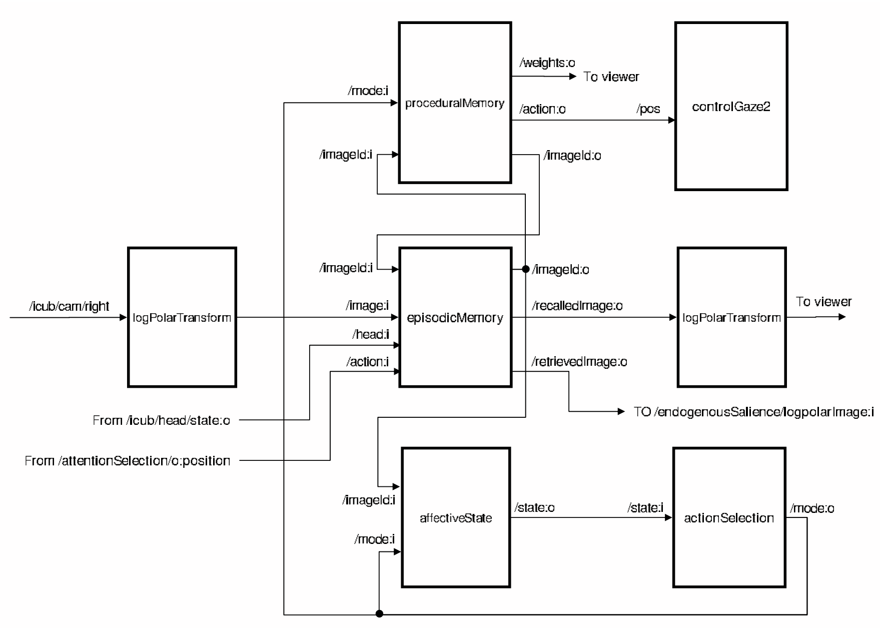


Fig. 7.6 Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part B)

7.3 The iCub Cognitive Architecture vs. the Roadmap Guidelines

As we noted above, the iCub cognitive architecture is a work-in-progress and what we have described is only a partial implementation of the roadmap guidelines set out in the previous chapter. We conclude the chapter by discussing how well the current version of the cognitive architecture follows these guidelines. Below, for each guideline, we will discuss what aspects have been followed and how they have been implemented, what aspects have not been followed, and how they might be. Again, we group the guidelines under seven headings, as we did in Chap. 6.[7]

7.3.1 Embodiment

Guideline 1 stipulates a rich array of physical sensory and motor interfaces. While the iCub has several extereoceptive sensors, including binocular vision, binaural hearing, a 3 degree-of-freedom vestibular sense, as well as the soon-to-be-deployed

[7] As noted above, Guidelines 1–9 are based directly on a study of enaction as a framework for development in cognitive robotics [385], as are the corresponding discussions in this section.

skin with cutaneous sensing, we consider only vision and hearing in the initial version of the iCub cognitive architecture. The remaining senses will be integrated at a later date. Concerning proprioceptive sensing, the iCub is specified with absolute position sensors on each joint for accurate servo-control. Force/torque sensors have been designed and are being deployed in the latest version of the robot but they are not yet considered in the cognitive architecture although they have been used in stand-alone mode to demonstrate compliant manipulation.

Guideline 3 stipulates a humanoid morphology. As we have seen in the first section of this chapter, the iCub follows this guideline faithfully, especially as the dimensions of the iCub are modelled on those of a human child and as it has such a high number of degrees of freedom, particularly in the hands and the head.

Guideline 13 requires that morphology should be integral to the model of cognition. This means that not only the robot's actions but also the robot's perceptions and the models it constructs of the world around it should depend on the physical form of the robot. It is clear that this follows naturally for action but what of perceptual processing and generative model construction? Two good examples of perception being dependent on the morphology of the robot are provided by Guidelines 28 & 29: the spatial and selective pre-motor theory of attention, whereby the location to which attraction is directed and the form to which is it attracted depends on the current configuration of the robot. However, neither of these pre-motor theories has yet been incorporated in the iCub cognitive architecture. On the other hand, a precursor to both of these theories is the ability to learn object affordances [118] (cf. Guidelines 7 & 41). Affordances can be modelled as associations between objects, action, and effects [249, 255, 256, 257]. This has been implemented on the iCub on a stand-alone basis [256, 257, 258] and it remains to integrate it into the iCub cognitive architecture in the internal simulation subsystem comprising the episodic memory and the procedural memory. We return to this issue in the next chapter.

7.3.2 Perception

Guideline 12 stipulates that attention should be fixated on the goal of an action. This means that there must be an attention system in operation (cf. Guideline 32) and that in the specific case of Guideline 12 the attention must be focussed on an anticipated future event which is the outcome of the action. This involves both spatial attention to direct the iCub gaze — where to look — and selective attention — what to look for. In the iCub cognitive architecture, prospection is effected by an internal simulation system which associates actions and perceptions. As we noted already, the internal simulation subsystem is effected by a combination of the episodic memory and the procedural memory. The attention system comprises endogenous salience, exogenous salience, egosphere, and attention selection. To direct attention selectively and spatially, the prospective internal simulation sub-system (specifically the episodic memory component) triggers the endogenous component.

Guideline 15 concerns the perception of regions in the optic array which possess the characteristics of objectedhood, viz.

- inner unity;
- persistent outer boundary;
- cohesive and distinct motion;
- relatively constant size and shape when in motion;
- a change in behaviour when in contact with other objects.

This capability entails several constituent processes (refer to Table 7.3). First, it is necessary to compute the instantaneous optical flow field. Second, its necessary to compensate for flow due to the movement of the iCub head; that is, it is necessary to compute visual motion with ego-motion compensation. Third, there needs to be a facility to group regions which exhibit distinct motion at a given instant and which also maintain some level of cohesion or constancy of this pattern of motion over a reasonable period of time. This grouping should be further strengthened by some measure of constancy of that region's boundary shape. Finally, this object perception capability should be modulated by a temporal pattern of actual and predicted gaze so that the perception of objecthood is strengthened or weakened, depending on whether or not these two patterns match one another. Although this capability is planned for the iCub, it has not yet been incorporated into the cognitive architecture, although doing so is a relatively straightforward matter.

Guideline 16 suggests that the system should develop the ability to discriminate between groups of one, two, and three objects but not necessarily higher numbers. The system should also be able to add small numbers up to a limit of three. Furthermore, it should also be able to discriminate between groups of larger numbers of objects provided that the ratio of the number of each group is large. Such a capability assumes the existence of a capability to perceived objects, in the sense of Guideline 15. It also requires an overt attention system with associated saccadic shifts in eye gaze and effective occular vergence to ensure the point of fixation, which is the perceived object, is coincident in both eyes. Both saccadic gaze control and occular vergence are currently included of the iCub cognitive architecture but the capability for number-dependent discrimination associated with Guideline 16 has not yet been implemented. It is envisaged that its implementation will be based on the discrimination between either overt or covert attentional scan path patterns, perhaps implemented as cliques in the procedural memory or a short-term variant of it. We will discuss the importance of short-term and long-term episodic and procedural memories, and the process of generalization that relates them, in the next chapter when we turn our attention to the challenges posed by completing the iCub cognitive architecture to embrace all forty-three guidelines.

Guideline 19 states that the iCub should be attracted to people and especially to their faces, their sounds, movements, and features. Again, this requires the presence of an attentional system, something we have already discussed and something that will arise several more times as we walk through these guidelines. As noted above, the iCub cognitive architecture does incorporate an attention system. In the specific case of Guideline 19, the attention system allows for spatial and selective attention to human faces and to relatively loud voices through binaural sound localization. However, this sound localization is not specific to the frequency spectrum of

Table 7.3 Constituent processes involved in the perceptuo-motor capabilities

Constituent Processes
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
Vergence
Gaze control: saccadic movement with prediction; possibly tuned by learning
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

humans; any loud noise will attract the iCub's attention. Similarly, attention to specific facial features such as eyes and mouth is envisaged but not yet implemented.

Guideline 20 elaborates further on the capabilities of the attentional system, specifying that attention should be preferentially attracted to biological motion. Again, this has not yet been implemented.

Guideline 21 deals with recognition of people, expressions, and action. Recognition on the iCub cognitive architecture is encapsulated to a degree in its episodic memory of past experiences (see Guideline 38). However, as currently implemented, this episodic memory is capable only of recognizing the instantaneous appearance of the current focus of attention. It doesn't distinguish between people and expressions and it doesn't yet incorporate any temporal or intentional component which would be necessary for action recognition. To an extent, action recognition is incorporated in the iCub procedural memory (see Guideline 39) which associates perceptions and actions. In this case, recognition can be effected by recalling the action associated with the perception prior to and following an action. This relates to Guidelines 7, 25, and 40 which are concerned with learning affordances, i.e. the association of possible action with perceived objects and consequent effects.

Guideline 22 returns again to the issue of attention, stipulating that gaze should be prolonged when a person looks directly at the iCub. This means that the attentional system should be attracted to a person's eyes and, in particular, to a pair of eyes that are looking directly at the iCub. Furthermore, it should have the capability to detect mutual gaze — when a person and the iCub make eye contact — and that this should inhibit temporarily other attentional mechanisms. At present, this capability is planned but has not yet been incorporated into the attention system.

Guideline 23 suggest that the iCub should receive and communicate emotion through facial gesture and that it should engage in turn-taking. While the iCub cognitive architecture does not have an ability to perceive the emotional state of someone with whom it is engaging, nor yet a capability for turn-taking, it can communicate its own emotional state, reflecting its current motivation (see Guidelines 6 and 34), through an elementary but engaging mechanism involving simulated eyebrows.

Guideline 26 states that the motor system should be involved in the semantic understanding of percepts, with procedural motor knowledge and internal action simulation being used to discriminate between percepts. Although there is no motoric component yet incorporated in the iCub episodic memory (Guideline 38), action and perception are mutually associated in the procedural memory and hence action can be incorporated into the recall of perceptions.

Guideline 27 suggest that the system should have a mechanism to learn hierarchical representations of regularities that can be then deployed to produce and perceive complex structured actions (i.e. intentional events which are not just simple sequences of physical states). This has not yet been incorporated into the cognitive architecture. Most probably, it will involve a significant extension of both episodic memory and procedural memory. We return to this important issue in the next chapter when we discuss some of the principal challenges posed by the roadmap guidelines.

Guidelines 28 and 29 deal with the pre-motor theory of spatial and selective attention whereby the preparation of a motor program in readiness to act in some spatial regions should predispose the perceptual system to process stimuli coming from that region and the preparation of a motor program in readiness to act on specific objects should predispose the perceptual system focus attention on those objects, respectively. Again, we see the significance of the attention system and, in this case, the motoric or proprioceptive modulation of attention. This capability has not yet been incorporated into the iCub cognitive architecture.

7.3.3 Action

Guideline 10 states that movements should be organized as actions. Since actions are planned, goal-directed, acts they are triggered by system motives (see Guideline 6) and they are guided by prospection. In other words, the action is defined, not by a servo-motor set-point specifying an effector movement, but by the goal of the action, an action whose outcome must be achieved adaptively by constituent movements. This guideline is implemented in the iCub cognitive architecture by the procedural and episodic memories and by use of visual servoing in reaching and locomotion. Specifically, the procedural memory represents actions as gaze saccades together with an optional reaching, hand-pushing, grasping, or locomotion movement. These actions are associated with expected outcomes defined by the expected outcome in the episodic memory. Thus, the procedural and episodic memory provide a feed-forward goal state for the action while the visual servoing, whereby the effector is adaptively controlled to align it with a fixation point while re-centering the gaze after the execution of the saccade, achieves the required motion through feedback control of the arm and hand.

Guideline 14 reflects the existence of pre-structured sensorimotor couplings at birth which serve to reduce the number of degrees of freedom of movement and thereby simplify the control problem. This pre-structuring is relaxed, allowing the number of degrees of freedom to increase, as development proceeds and as the associated skill is mastered. At present, the iCub cognitive architecture does not implement this guideline. To do so would require a more sophisticated learning strategy in, for example, the reach and grasp component.

Guideline 17 stipulates that navigation should be based on representations that are dynamic and ego-centric so that navigation is based on path integration by moving from place to place using local landmarks. The iCub cognitive architecture implements this guideline directly through its episodic and procedural memoryies. The episodic memory holds the landmark appearances while the procedural memory represents the locomotion action, A_j required to take the iCub from one landmark P_i to another P_k. Navigation is effected by traversing some path through this procedural memory, taking the iCub from some initial landmark P_a to some final landmark P_z via intermediate landmarks, viz. $P_a, A_1, P_b, A_2, P_c, \ldots, P_z$.

Guideline 18 is related to Guideline 17 and stipulates that re-orientation should be effected by recognizing landmarks and not by a global representation of the environment. This is exactly what is done in the iCub procedural memory and episodic memory. Guideline 18 goes on to stipulate that it is the view-dependence of the landmark that is important for the re-orientation so that the geometry of the landmark rather than distinctive features that is used. This part of the guideline has not been implemented. Since the episodic memory currently uses colour histograms of log-polar images, acquired when the iCub is fixating on some point of interest such as a landmark, the landmark recognition is based on scale-invariant appearance rather than local geometry.

Guideline 36 states that action selection should be modulated by affective motivation mechanisms. The iCub cognitive architecture follow this guideline directly, albeit in a very simple form at present since the behaviour of the action selection module is dependent only on the output of the affective state module.

Guideline 42, which suggests that the system should have hierarchically-structured representations for the acquisition, decomposition, and execution of action-sequence skills, has not yet been implemented in the cognitive architecture either. We discuss this issue further in the next chapter.

7.3.4 Anticipation

Guidelines 8 and 35 — which state that the system should incorporate processes of internal simulation to scaffold knowledge and to facilitate prediction of future events, explanation of observed events, and the imagination of new events — is implemented directly in the cognitive architecture through procedural memory with prediction being effected by following a sequence of perception-action-perception associations $P, A, P, A, \ldots$ forward in time along the path with the strongest associative connections. Explanation (or reconstruction) follows the path backward in time and imagination follows it forward along a path with weak associative connections.

However, the lack of a capacity for generalization limits the power of this internal simulation at present.

7.3.5 Adaptation

Guideline 4 goes to the heart of development. It stipulates that the system must support developmental processes that modify the system's structure so that its dynamics of interaction are altered to effect an increase in the space of viable actions and an extension of the time horizon of the system's anticipatory capability. The internal simulation system comprising the episodic and procedural memories in the iCub cognitive architecture accomplishes this to a limited extent. As the iCub explores its environment, as it looks around, guided by attentive processes that are triggered by both internal and external stimuli, as it moves, reaches, grasps, manipulates, it learns to associate perceptions with actions and actions with perceptions and thereby develops an understanding of its environment which as we have seen can then be used to predict and act. However, it is a weak form of development: it learns from experience how things are, rather than how things might be. That is, the current iCub cognitive architecture has no capacity for generalization. While recurrent action-perception associations are indeed strengthened by exploration and experience, there is no generative mechanism which constructs models of these action-perception associations that go beyond these particular instances to capture a more encompassing lawfulness in the iCub's interactions. Put another way, the iCub's cognitive architecture currently has no way of building a model by extrapolating from experience and then validating, refining, or discarding that extrapolated model. We return to this crucial area in the next chapter.

Guideline 5 builds on Guideline 4 by requiring that the system should operate autonomously so that developmental changes are not a deterministic reaction to an external stimulus but result from an internal process of generative model construction. Notwithstanding the fact that the present iCub cognitive architecture implements Guideline 4 in a weak manner, the mechanism which governs the construction of these procedural models are nontheless autonomous: they depend only on the affective state of the system which depend in turn on how well the outcome of its explorative actions match its expectations. The system's goals are driven entirely by the internal affective processes.

Guideline 7 requires the ability to learn object affordances [118]. As we noted above in Sect. 7.3.1, affordances can be modelled as associations between objects, action, and effects [249, 255, 256, 257]. This has been implemented on the iCub on a stand-alone basis [256, 257, 258] and it remains to integrate it into the iCub cognitive architecture in the internal simulation subsystem comprising the episodic memory and the procedural memory.

Guideline 9 says that the system should also incorporate processes for grounding internal simulations in actions to establish by observation their validity. This is accomplished in the iCub cognitive architecture by the Affective State, Action Selection, Endogenous Salience, and Episodic Memory. Specifically, when the affective state is in an explorative state, the endogenous salience is primed by an episodic memory representing the expected outcome of an action which is about to be performed. If the subsequently acquired percept matches the expectation, then the perception-action association is strengthened. If it isn't then the affective state changes from exploration to curiousity and is driven by exogenous factors, not internally-generated endogenous ones.

Guideline 33 states the system should learn from experience the motor skills associated with actions. This is implemented directly in the iCub cognitive architecture in the reach component which learns through trial and error how to place a hand at the point in the field of view where it currently fixating. Since actions are specified in part by gaze fixation, the motor skills associated with action are indeed learning from experience.

Guideline 38 stipulates that the system should have both transient and generalized episodic memory of past experiences. As we have noted above, the iCub cognitive architecture has a transient episodic memory but not yet the capacity to generalize. Again, we return to this important issue in the next chapter.

Guideline 39 says that the system should have a procedural memory of actions and outcomes associated with episodic memories. This guideline has been followed faithfully in the design of the iCub cognitive architecture.

7.3.6 Motivation

Guideline 6 stipulates that development must be driven by internally-generated social and exploratory motives which enable the discovery of novelty and regularities in the world and the potential of the system's own actions. This guideline has been partially followed in that explorative motives have been implemented in the affective state but as yet social motives have not. Two forms of explorative motive have been implemented: curiousity and experimentation, focussing on exogenous and endogenous events respectively. It is envisaged that social motives will balance the two, the main idea being that in social interaction, a cognitive agent is trying to establish a common epistemology with the social partners and this requires equal attention to interactions generated by the partner (which have to be assimilated into the model the agent is constructing) and interactions generated by the agent (which are attempts to ground that model by interacting with the partner to see if the agent's expression of that model in the interaction is understood by the partner).

Guideline 34 states that the system should have a spectrum of self-regulating autonomy-preserving homeostatic processes associated with different levels of emotion or affect resulting in different levels of cognitive function and behavioural complexity. At present, there is just one very simple homeostatic process in the action selection component. Clearly this needs to be remedied in the future.

7.3.7 Autonomy

The perceptuo-motor capabilities outlined in the previous sections operate concurrently, competing for control. A cognitive architecture must provide a mechanism

Table 7.4 Guidelines for the configuration of the phylogeny of a humanoid robot vis-à-vis the degree of adoption in the iCub cognitive architecture. Key: '×' indicates that the guideline has been strongly implemented in the architecture, '+' indicates that it is weakly or minimally implemented, and a space indicates that it has not yet been followed in any substantial manner. Similar guidelines derived from more than one source (i.e. from Enaction, Developmental Psychology, Neurophysiology, or Computational Modelling) have been combined. Secondary source guidelines are shown in brackets.

Guidelines for the Phylogeny of a Developmental Cognitive System		
Number	**Guideline**	**iCub**
	Embodiment	
1	Rich array of physical sensory and motor interfaces	+
3	Humanoid morphology	×
13	Morphology integral to the model of cognition	+
	Perception	
12 (32)	Attention fi xated on the goal of an action	×
15	Perception of objecthood	
16	Discrimination & addition of small numbers; groups of large numbers	
19	Attraction to people (faces, their sounds, movements, and features)	+
20	Preferential attention to biological motion	
21	Recognition of people, expression, and action	+
22	Prolonged attention when a person engages in mutual gaze	
23	Perceive & communicate emotions by facial gesture and engage in turn-taking	+
26	Involvement of the motor system in discrimination between percepts	+
27	Mechanism to learn hierarchical representations	
28	Pre-motor theory of attention —spatial attention	
29	Pre-motor theory of attention —selective attention	
	Action	
10	Movements organized as actions	×
14	Early movements constrained to reduce the number of degrees of freedom	
17	Navigation based on dynamic ego-centric path integration	×
18	Re-orientation based on local landmarks	+
36	Action selection modulated by affective motivation mechanisms	×
42	Hierarchically-structured representations of action-sequence skills	
	Anticipation	
8, 35	Internal simulation to predict, explain, & imagine events, and scaffold knowledge	×
	Adaptation	
4	Self-modifi cation to expand actions and improve prediction	+
5	Autonomous generative model construction	+
7 (25, 41)	Learning affordances	×
9 (40)	Grounding internal simulations in actions	×
33	Learn from experience the motor skills associated with actions	×
38	Transient and generalized episodic memories of past experiences	+
39	Procedural memory of actions and outcomes associated with episodic memories	×
	Motivation	
6 (11, 31)	Social and explorative motives	+
34	Affective drives associated with autonomy-preserving processes of homeostasis	+
	Autonomy	
2	Autonomy-preserving processes of homeostasis	+
24	Encode space in motor & goal specifi c manner	+
30	Minimal set of innate behaviours for exploration and survival	+
37	Separate representations associated with each component / sub-system	+
43	Concurrent competitive operation of components and subsystems	×

for modulating or deploying these phylogenetic perceptuo-motor capabilities: for selecting from among the many potential actions that are competing for realization. This is exactly a homeostatic process which acts to preserve the autonomy of the iCub while fulfilling it cognitive drive to increase its space of action and improve its anticipatory capability.

Guideline 2 states that system should exhibit structural determination: that is, the system should have a range of autonomy-preserving processes of homeostasis that maintain the system's operational identity and thereby determine the meaning of the system's interactions. As noted above, at present, there is just one very simple homeostatic process in the action selection component.

Guideline 24 states that the system should encode space in several different ways, each of which is specifically concerned with a particular motor goal. At present, space is encoded in only one way by eye gaze which in turn is used to direct other actions such as reaching and locomotion. The two most obvious candidates for a second encoding of space in the iCub cognitive architecture is through proprioceptive grasp configuration aided by finger sensor data, and through a sense of peripersonal space mediated by artificial skin covering the body explored by the iCub's hands.

Guideline 30 stipulates that the system should have a minimal set of innate behaviours for exploration and survival, i.e. preservation of autonomy. The iCub cognitive architecture has at present several innate behaviours such as attention-directed gaze control, gaze-controlled reach and grasp, gaze-controlled locomotion, and affective action selection.

Guideline 37 states that the system should have separate and limited representations of the world and the task at hand in each component / sub-system. This guideline is followed in the iCub cognitive architecture in the sense that there is no global representation shared by all sub-systems. Each sub-system operates as a distinct module with its own encapsulated representations.

Guideline 43 stipulates that the components / sub-systems of the cognitive architecture should operate concurrently so that the resultant behaviour emerges as a sequence of states arising from their interaction as they compete. This is the very essence of the iCub cognitive architecture. It is made possible by YARP and it results in an emergent behaviour as a sequence of perception, action, and cognitive states that are the result of internal dynamics of interaction between these components and not some pre-determined state machine.

7.4 Summary

Table 7.4 summarizes the degree to which the roadmap guidelines have been adopted in the iCub cognitive architecture. Some guidelines have been strongly implemented (incidated by a '$\times$' symbol in the table), some have been weakly or minimally implemented (indicated by a '$+$'), and other have not yet been followed in any substantial manner (indicated by a space).

Chapter 8
Conclusion

Drawing on the insights from Chaps. 1 to 5, Chap. 6 presented the core of this book: a comprehensive list of forty-three guidelines for the design of an enactive cognitive architecture and its practical deployment as a roadmap of cognitive development in a humanoid robot. Chap. 7 discussed in detail how these guidelines were used to influence the design and implementation of a cognitive architecture for the iCub humanoid robot. We saw that, although many of the guidelines were followed, several were either only partly followed and some have not yet been followed at all (see Table 7.4 in the previous chapter). We emphasize here that these omissions are not because these guidelines are not important — quite the opposite — but because the iCub cognitive architecture, like all cognitive architectures, is a work-in-progress and future versions will reflect more complete implementation of all guidelines. In accomplishing this, we will inevitably face some significant challenges and, in this chapter, we wish to bring the book to a close by re-visiting some of the issues that are particularly pivotal to cognitive development, in general, and the complete implementation of the forty-three guidelines in the iCub cognitive architecture, in particular.

8.1 Multiple Mechanisms for Anticipation

We have emphasized in this book the importance of anticipation in cognition and we focussed in particular on the use of internal simulation to achieve this prospection. However, prospection and anticipation may require different approaches in different circumstances. Ideally, they should be accomplished in several complementary ways as there is often a need to allow direct anticipatory control of perceptuo-motor abilities without having to revert to 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 more temporally-extended prospection circuit based on internal simulation.

D. Vernon et al.: A Roadmap for Cogn. Develop. in Humanoid Robots, COSMOS 11, pp. 155–158.
springerlink.com

8.2 Prediction, Reconstruction, and Action: Learning Affordances

Every action entails a prediction about how the perceptual world will change as a consequence of that action. This is the goal of the action and it is what differentiates an action from a simple movement or sequence of movements. 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 simple prediction, reconstruction (or explanation), and action as associative recall by presenting $(P_i, A, \sim)$, $(\sim, A, P_j)$, or $(P_i, \sim, P_j)$, respectively, to the iCub's Procedural Memory and by associatively recalling the missing element. This triplet-based representation, is conceptually identical to the stand-alone iCub framework for learning object affordances [256, 257, 258]. What differs is the manner in which the association network is constructed. In the latter framework, learning is accomplished by autonomous exploration using elementary actions that allow the iCub to experiment with its environment and to develop and understanding the relationships between actions, objects, and action outcomes, modelling these relationships using a Bayesian network. 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. $(O, A) \rightarrow E$ is the predictive aspect of affordance; $(O, E) \rightarrow A$ recognizes an action and aids planning; $(A, E) \rightarrow O$ is object recognition and selection. One of our immediate goals is to integrate this affordance learning technique with the Procedural Memory and Episodic Memory components in the iCub cognitive architecture.

A significant problem remains, however. In the existing framework, the actions that the robot uses to experiment with and explore the object are assumed to exist as predefined primitive manipulation movements, such as push, tap, and grasp. Clearly, we require a more flexible approach in which the action's movements can be generated and adjusted adaptively as a consequence of the outcome of the action.

8.3 Object Representation

At present, there is no explicit concept of objecthood in the iCub cognitive architecture in the sense of Guideline 15, and there are no visual processes which identifies such objects. As we noted in the previous chapter, the implementation of this capability is far from trivial but, in principle, it doesn't pose a major challenge using conventional computer vision techniques based on motion segmentation and boundary detection. However, it would be instructive to investigate a more active approach based on the characteristics of overt attention, since attention mirrors interpretation and it plays such a significant part in cognition. Arguably, and consistent with Guideline 15, 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 $(P_a, A_i, P_b \ldots A_j, P_c)$ clique within the network of associations in the procedural memory. Object detection and recognition then becomes a matter of associative clique retrieval based one all or part of the clique.

Appendix A
Catalogue of Cognitive Architectures

This appendix contains a catalogue of twenty cognitive architectures drawn from the cognitivist, emergent, and hybrid traditions, beginning with some of the best known cognitivist ones. Table A.1 lists the cognitive architectures surveyed.

Table A.1 The cognitive architectures surveyed in this appendix. This survey is adapted from [387] and extended to bring it up to date by including the GLAIR, CoSy Architecture Schema, Cognitive-Affective, LIDA, CLARION, PACO-PLUS cognitive architectures. The architectures are treated in order, top-to-bottom and left-to-right (*i.e.* cognitivist first, then emergent, and finally hybrid).

Cognitivist	**Emergent**	**Hybrid**
Soar	AAR	HUMANOID
EPIC	Global Workspace	Cerebus
ACT-R	I-C SDAL	Cog: Theory of Mind
ICARUS	SASE	Kismet
ADAPT	DARWIN	LIDA
GLAIR	Cognitive-Affective	CLARION
CoSy		PACO-PLUS

A.1 Cognitivist Cognitive Architectures

A.1.1 The Soar Cognitive Architecture

The Soar system [211, 326, 220, 222] is Newell's candidate for a Unified Theory of Cognition [271] and, as such, it is an architypal cognitivist cognitive architecture (as well as being an iconic one).It is a production (or rule-based) system[1] that operates in a cyclic manner, with a production cycle and a decision cycle. It operates as follows. First, all productions that match the contents of declarative (working) memory fire. A production that fires may alter the state of declarative memory and cause other productions to fire. This continues until no more productions fire. At this point, the decision cycle begins in which a single action from several possible actions is selected. The selection is based on stored action preferences. Thus, for each decision cycle there may have been many production cycles. Productions in Soar are low-level; that is to say, knowledge is encapsulated at a very small grain size.

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

[1] 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.

References

1. Adolph, K.E., Eppler, M.A.: Development of visually guided locomotion. Ecological Psychology 10, 303–321 (1998)
2. Aguiar, A., Baillargeon, R.: Developments in young infants' reasoning about occluded objects. Cognitive Psychology 45, 267–336 (2002)
3. Aleksander, I.: Neural systems engineering: towards a unified design discipline? Computing and Control Engineering Journal 1(6), 259–265 (1990)
4. Andersen, R.A., Asanuma, C., Essick, G., Siegel, R.M.: Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. J. Comp. Neurol. 296, 65–113 (1990)
5. Andersen, R.A., Gnadt, J.W.: Role of posterior parietal cortex. In: Wurtz, R.H., Goldberg, M.E. (eds.) The Neurobiology of Saccadic Eye Movements, Reviews of Oculomotor Research, vol. 3, pp. 315–335. Elsevier, Amsterdam (1989)
6. Anderson, J.R.: Act: A simple theory of complex cognition. American Psychologist 51, 355–365 (1996)
7. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111(4), 1036–1060 (2004)
8. Anokhin, P.K.: Systemogenesis as a general regulator of brain development. Progress in Brain Research 9 (1964)
9. Apkarian, P.: Temporal frequency responsivity shows multiple maturational phases: State-dependent visual evoked potential luminance flicker fusion from birth to 9 months. Visual Neuroscience 10(6), 1007–1018 (1993)
10. Arbib, M.A.: Perceptual structures and distributed motor control. In: Brooks, V.B. (ed.) Handbook of Physiology, Section 1: The Nervous System, vol. II: Motor Control, pp. 1449–1480. Williams and Wilkins, Baltimore (1981)
11. Arkin, A.: Behavior-based Robotics. MIT Press, Cambridge (1998)
12. Aslin, R.N.: Development of smooth pursuit in human infants. In: Fisher, D.F., Monty, R.A., Senders, J.W. (eds.) Eye Movements: Cognition and Visual Perception. Erlbaum, Hillsdale (1981)
13. Aslin, R.N., Shea, S.L.: Velocity thresholds in human infants: Implications for the perception of motion. Developmental Psychology 26, 589–598 (1990)
14. Atkinson, J.: The developing visual brain. Oxford University Press, Oxford (2000)
15. Baars, B.J.: A Cognitive Theory of Consciousness. Cambridge University Press, Cambridge (1998)
16. Baars, B.J.: The conscious assess hypothesis: origins and recent evidence. Trends in Cognitive Science 6(1), 47–52 (2002)
17. Baars, B.J., Franklin, S.: Consciousness is computational: The LIDA model of global workspace theory. International Journal of Machine Consciousness 1(1), 23–32 (2009)
18. Badaley, D., Adolph, K.: Beyond the average: Walking infants take steps longer than their leg length. Infant Behavior and Development 31, 554–558 (2008)
19. Baillargeon, R., Graber, M.: Where's the rabbit? 5.5-month-old infants' representation of the height of a hidden object. Cognitive Development 2, 375–392 (1987)
20. Banks, M.S., Bennett, P.J.: Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. Journal of the Optical Society of America 5(12), 2059–2079 (1988)
21. Barbu-Roth, M., Anderson, D., Despres, A., Provasi, J., Campos, J.: Neonatal stepping to terrestial optic flow. Child Development 80(1), 8–14 (2009)

22. Barela, J.A., Godoia, D., Freitas, P.B., Polastria, P.F.: isual information and body sway coupling in infants during sitting acquisition. Infant Behavior and Development 23(3-4), 285–297 (2000)
23. Barela, J.A., Jeka, J.J., Clark, J.E.: The use of somatosensory information during the aquisition of independent upright stance. Infant Behavior and Development 22(1), 87–102 (1999)
24. Barnes, G.R.: 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)
25. Baron-Cohen, S.: Mindblindness. MIT Press, Cambridge (1995)
26. Barrett, T., Needham, A.: Developmental differences in infants use of an objects shape to grasp it securely. Developmental Psychobiology 50, 97–106 (2009)
27. Barth, H., Kanwisher, N., Spelke, E.: The construction of large number representations in adults. Cognition 86, 201–221 (2003)
28. Bates, E., Camaioni, L., Volterra, V.: The acquisition of preformatives prior to speech. Merril-Palmer Quarterly 21(3), 205–226 (1975)
29. Bayley, N.: Bayley scales of infant development. Psychological Corporation, New York (1969)
30. Benjamin, D., Lyons, D., Lonsdale, D.: Adapt: A cognitive architecture for robotics. In: Hanson, A.R., Riseman, E.M. (eds.) 2004 International Conference on Cognitive Modeling, Pittsburgh, PA (2004)
31. Benson, A.J., Barnes, G.R.: Vision during angular oscillations: The dynamic interaction of visual and vestibular mechanisms. Aviation Space and Environmental Medicine 49, 340–345 (1978)
32. Bernstein, N.: The coordination and regulation of movements. Pergamon, Oxford (1967)
33. Bertenthal, B., Gredebäck, G.: Information signifying occlusion in infants (2006) (in preparation)
34. Bertenthal, B., Rose, J., Bai, D.: Perception-action coupling in the development of visual control of posture. Journal of Experimental Psychology: Human Perception and Performance 23, 1631–1643 (1997)
35. Berthier, N., Clifton, R.K., Gullapalli, V., McCall, D.D., Robin, D.J.: Visual information and object size in the control of reaching. J. Mot. Behav. 28(3), 187–197 (1996)
36. Berthier, N.E.: Learning to reach: a mathematical model. Developmental Psychology 32, 811–823 (1996)
37. Berthoz, A.: The Brain's Sense of Movement. Harvard University Press, Cambridge (2000)
38. Berton, F.: A brief introduction to log-polar mapping. In: Technical Report, LIRA-Lab. University of Genova (2006)
39. Berton, F., Sandini, G., Metta, G.: Anthropomorphic visual sensors. In: Encyclopedia of Sensors, vol. 10, pp. 1–16 (2006)
40. Bickhard, M.H.: Autonomy, function, and representation. Artificial Intelligence, Special Issue on Communication and Cognition 17(3-4), 111–131 (2000)
41. Bienenstock, E.L., Cooper, L.N., Munro, P.W.: Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. Journal of Neurscience 2(1), 32–48 (1982)
42. Birch, E.E., Gwiazda, J., Held, R.: Stereoacuity development for crossed and uncrossed disparities in human infants. Vision Research 22, 507–513 (1982)
43. Bloch, H., Carchon, I.: On the onset of eye-head co-ordination in infants. Behavioural Brain Research 49, 85–90 (1992)

44. Braccini, C., Gambardella, G., Sandini, G.: A signal theory approach to the space and frequency variant filtering performed by the human visual system. Signal Processing 3, 231–240 (1981)
45. Breazeal, C.: Sociable Machines: Expressive Social Exchange Between Humans and Robots. Unpublished Doctoral Dissertation. MIT, Cambridge (2000)
46. Breazeal, C.: Emotion and sociable humanoid robots. International Journal of Human-Computer Studies 59, 119–155 (2003)
47. Breazeal, C., Scassellati, B.: A context-dependent attention system for a social robot. In: Proc. of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 1999, Stockholm, Sweden, pp. 1146–1151 (1999)
48. Brodmann, K.: Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargerstellt auf Grund des Zellenbaues. Barth, Leipzig (1909)
49. Brooks, R.A.: A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation RA-2(1), 14–23 (1986)
50. Brooks, R.A.: Flesh and Machines: How Robots Will Change Us. Pantheon Books, New York (2002)
51. Brooks, R.A., Breazeal, C., Marajanovic, M., Scassellati, B., Williamson, M.M.: The cog project: Building a humanoid robot. In: Nehaniv, C.L. (ed.) CMAA 1998. LNCS (LNAI), vol. 1562, p. 52. Springer, Heidelberg (1999)
52. Brotchie, P.R., Andersen, R.A., Snyder, L.H., Goodman, S.J.: Head position signals used by parietal neurons to encode locations of visual stimuli. Nature 375, 232–235 (1995)
53. Bruce, C.J., Goldberg, M.E.: Primate frontal eye fields. I. single neurons discharging before saccades. J. Neurophysiol. 53, 603–635 (1985)
54. Burghart, C., Mikut, R., Stiefelhagen, R., Asfour, T., Holzapfel, H., Steinhaus, P., Dillman, R.: A cognitive architecture for a humanoid robot: A first approach. In: IEEE-RAS International Conference on Humanoid Robots (Humanoids 2005), pp. 357–362 (2005)
55. Byrne, M.D.: Cognitive architecture. In: Jacko, J., Sears, A. (eds.) The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications, pp. 97–117. Lawrence Erlbaum, Mahwah (2003)
56. Camaioni, L., Caselli, M.C., Longbardi, E., Volterra, V.: A parent report instrument for early language assessment. First Language 11, 345–360 (1991)
57. Carpenter, M., Nagell, K., Tomasello, M.: Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monogr. Soc. Res. Child Dev. 63(4), 1–143 (1998)
58. Carré, R., Lindblom, B., MacNeilage, P.: Rôle de l'acoustique dans l'évolution du conduitvocal humain. C. R. Acad. Sci. Paris Tome 320(Srie IIb) (1995)
59. Cavada, C., Goldman-Rakic, P.S.: 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)
60. Choi, D., Kaufman, M., Langley, P., Nejati, N., Shapiro, D.: An architecture for persistent reactive behavior. In: Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 988–995. ACM Press, New York (2004)
61. Christensen, W.D., Hooker, C.A.: An interactivist-constructivist approach to intelligence: self-directed anticipative learning. Philosophical Psychology 13(1), 5–45 (2000)
62. Christensen, W.D., Hooker, C.A.: Representation and the meaning of life. In: Representation in Mind: New Approaches to Mental Representation, The University of Sydney (2000)
63. Clark, A.: Mindware – An Introduction to the Philosophy of Cognitive Science. Oxford University Press, New York (2001)

64. Clark, H.H.: Managing problems in speaking. Speech Communication 15, 243–250 (1994)
65. Clifton, R.K., Muir, D.W., Ashmead, D.H., Clarkson, M.G.: Is visually guided reaching in early infancy a myth? Child Development 64(4), 1099–1110 (1993)
66. Corbetta, D., Thelen, E.: The developmental origins of bimanual coordination: a dyamic perspective. J. Exp. Psychol. Hum. Percept. Perform. 22(2), 502–522 (1996)
67. Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E.: Attentional modulation of neural processing of shape, color, and velocity in humans. Science 248, 1556–1559 (1990)
68. Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E.: Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J. Neurosci. 11, 2383–2402 (1991)
69. Corkum, V., Moore, C.: The origins of joint visual attention in infants. Developmental Psychology 34, 28–38 (1998)
70. Craighero, L., Fadiga, L., Rizzolatti, G., Umilta, C.A.: Movement for perception: a motor-visual attentional effect. Journal of Experimenal Psychology: Human Perception and Performance (1999)
71. Craighero, L., Nascimben, M., Fadiga, L.: Eye position affects orienting of visuospatial attention. Current Biology 14, 331–333 (2004)
72. Crutchfield, J.P.: Dynamical embodiment of computation in cognitive processes. Behavioural and Brain Sciences 21(5), 635–637 (1998)
73. Damasio, A.R.: Looking for Spinoza: Joy, sorrow and the feeling brain. Harcourt, Orlando (2003)
74. Dannemiller, J.L., Freedland, R.L.: The detection of slow stimulus movement in 2- to 5-month-old infants. Journal of Experimental Child Psychology 47, 337–355 (1989)
75. Dayton, G.O., Jones, M.H.: Analysis of characteristics of fixation reflex in infants by use of direct current electro-oculography. Neurology 14, 1152–1156 (1964)
76. Deák, G.O., Flom, R.A., Pick, A.D.: 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)
77. deCasper, A.J., Fifer, W.P.: On human bonding: Newborns prefer their mothers' voices. Science 208, 1174–1176 (1980)
78. Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, B., Mazziotta, J.C.: Mapping motor representations with PET. Nature 371, 600–602 (1994)
79. Degallier, S., Righetti, L., Ijspeert, A.J.: Hand placement during quadruped locomotion in a humanoid robot: A dynamical system approach. In: IROS, San Diego, USA (2007)
80. D'Entremont, B., Hains, S.M.J., Muir, D.W.: A demonstration of gaze following in 3- to 6-month-olds. Infant Behavior and Development 20, 569–572 (1997)
81. Elk, M.V., van Schie, H.T., Hunnius, S., Vesper, C., Bekkering, H.: You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. Neuroimage 43(4), 808–814 (2008)
82. Ojemann, G., et al.: Cortical language localization in left, dominant hemisphere: an electrical stimulation mapping investigation in 117 patients. J. Neurosurg. 71, 316–326 (1989)
83. Schaffler, L., et al.: Comprehensive deficits elecited by electrical stimulation of broca's area. Brain 116, 695–715 (1993)
84. Fadiga, L., Craighero, L., D'Ausilio, A.: Broca's area in language, action, and music. The Neurosciences and Music III — Disorders and Plasticity: Ann. N. Y. Acad. Sci. 1169, 448–458 (2009)

85. Fadiga, L., Fogassi, L., Pavesi, G., Rizzolatti, G.: Motor facilitation during action observation: a magnetic stimulation study. Journal of Neurophysiology 73(6), 2608–2611 (1995)
86. Fadiga, L., Gallese, V.: Action representation and language in the brain. Theoretical Linguistics 23, 267–280 (1997)
87. Fagard, J., Lockman, J.J.: The effect of task constraints on infants (bi)manual strategy for grasping and exploring objects. Infant Behavior & Development 28, 305–315 (2005)
88. Fagard, J., Spelke, E.S., von Hofsten, C.: Reaching and grasping a moving object in 6-, 8-, and 10-month olds: laterality aspects (2008) (submitted manuscript)
89. Farroni, T., Csibra, G., Simeon, F., Johnson, M.H.: Eye contact detection in humans from birth. PNAS 99, 9602–9605 (2002)
90. Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A.C., Pozzo, T., Calzolari, F., Granieri, E., Fadiga, L.: Encoding of human action in broca's area. Brain (2009)
91. Feigenson, L., Carey, S., Hauser, M.: The representations underlying infants choice of more: object-files versus analog magnitudes. Psychological Science 13, 150–156 (2002)
92. Feigenson, L., Dehaene, S., Spelke, E.S.: Core systems of number. Trends in Cognitive Sciences 8, 307–314 (2004)
93. Feigenson, L., et al.: The representations underlying infants' choice of more: object-files versus analog magnitudes. Psychological Science 13, 150–156 (2002)
94. Fiebach, C.J., Vos, S.H., Friederici, A.D.: Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. J. Cog. Neurosci. 16, 1562–1575 (2004)
95. Fikes, R.E., Nilsson, J.J.: Strips: A new approach to the application of theorem proving to problem solving. Artificial Intelligence 2, 189–208 (1971)
96. Fitzpatrick, P., Metta, G., Natale, L.: Towards long-lived robot genes. Journal of Robotics and Autonomous Systems 56, 1–3 (2008)
97. Flanagan, J.R., Johansson, R.S.: Action plans used in action observation. Nature 424, 769–771 (2003)
98. Fodor, J.A.: Modularity of Mind: An Essay on Faculty Psychology. MIT Press, Cambridge (1983)
99. Fodor, J.A.: The Mind Doesn't Work that Way. MIT Press, Cambridge (2000)
100. Forssberg, H.: Ontogeny of human locomotor control. i. infant stepping, supported locomotion, and transition to independent locomotion. Experimental Brain Research 57, 480–493 (1985)
101. Fox, R., McDaniel, C.: The perception of biological motion by human infants. Science 218, 486–487 (1982)
102. Fraiberg, S.: Insights from the blind. Basic Books, New York (1977)
103. Franklin, S.: A foundational architecture for artificial general intelligence. In: Goertzel, B., Wang, P. (eds.) Proceeding of the 2007 Conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms, pp. 36–54. IOS Press, Amsterdam (2007)
104. Franklin, S., Ramamurthy, U., D'Mello, S.K., McCarthy, L., Negatu, A., Silva, R., Datla, V.: LIDA: A computational model of global workspace theory and developmental learning. In: AAAI Fall Symposium on AI and Consciousness: Theoretical Foundations, pp. 61–66 (2007)
105. Freeman, W.J., Núñez, R.: Restoring to cognition the forgotten primacy of action, intention and emotion. Journal of Consciousness Studies 6(11-12), ix–xix (1999)
106. Friedlander, D., Franklin, S.: LIDA and a theory of mind. In: Proceeding of the 2008 Conference on Advances in Artificial General Intelligence, pp. 137–148. IOS Press, Amsterdam (2008)

107. Frintrop, S.: VOCUS: A Visual Attention System for Object Detection and Goal-directed Search. Rheinische Friedrich-Wilhelms-Universitat Bonn Institut Fur Informatik and Freunhofer Institut fur Autonome Intelligente Systeme (2006); Ph.D. Thesis
108. Froese, T., Ziemke, T.: Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artificial Intelligence 173, 466–500 (2009)
109. Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G.: Action recognition in the premotor cortex. Brain 119, 593–609 (1996)
110. Gallistel, C.R.: The organization of learning. MIT Press, Cambridge (1990)
111. Gardner, H.: Multiple Intelligences: The Theory in Practice. Basic Books, New York (1993)
112. Gentilucci, M., Daprati, E., Gangitano, M.: Implicit visual analysis in handedness recognition. Consciousness & Cognition 7, 478–493 (1998)
113. Gentilucci, M., Daprati, E., Gangitano, M.: Right-handers and left-handers have different representations of their own hand. Cognitive Brain Research 6, 185–192 (1998)
114. Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Camarda, R., Rizzolatti, G.: Functional organization of inferior area 6 in the macaque monkey. I. somatotopy and the control of proximal movements. Exp. Brain Res. 71, 475–490 (1988)
115. Gentilucci, M., Scandolara, C., Pigarev, I.N., Rizzolatti, G.: Visual responses in the postarcuate cortex (area 6) of the monkey that are independent of eye position. Exp. Brain Res. 50, 464–468 (1983)
116. Gibson, E.J., Pick, A.: An Ecological Approach to Perceptual Learning and Development. Oxford University Press, Oxford (2000)
117. Gibson, J.J.: The senses considered as perceptual systems. Houghton Mifflin, New York (1966)
118. Gibson, J.J.: The theory of affordances. In: Shaw, R., Bransford, J. (eds.) Perceiving, Acting and Knowing: Toward an Ecological Psychology, pp. 67–82. Lawrence Erlbaum, Mahwah (1977)
119. Gillner, S., Mallot, H.A.: Navigation and acquisition of spatial knowledge in a virtual maze. Journal of Cognitive Neuroscience 10, 445–463 (1998)
120. Goldberg, M.E., Segraves, M.A.: The visual and frontal cortices. in: The neurobiology of saccadic eye movements. In: Wurtz, R.H., Goldberg, M.E. (eds.) Reviews of Oculomotor Research, vol. 3, pp. 283–313. Elsevier, Amsterdam (1989)
121. Goldin-Meadow, S., Butcher, C.: Pointing towards two-word speech in young children. In: Kita, S. (ed.) Pointing: Where Language, Culture and Cognition Meet, pp. 85–187. Erlbaum, Mahwah (2003)
122. Grafton, S.T., Arbib, M.A., Fadiga, L., Rizzolatti, G.: Localization of grasp representations in humans by PET: 2. observation compared with imagination. Experimental Brain Research 112, 103–111 (1996)
123. Granlund, G.: Organization of architectures for cognitive vision systems. In: Christensen, H.I., Nagel, H.-H. (eds.) Cognitive Vision Systems. LNCS, vol. 3948, pp. 39–58. Springer, Heidelberg (2005)
124. Granlund, G.H.: The complexity of vision. Signal Processing 74, 101–126 (1999)
125. Granlund, G.H.: Does vision inevitably have to be active? In: Proceedings of SCIA 1999, Scandanavian Conference on Image Analysis (1999); Also as Linkping University Technical Report LiTH-ISY-R-2247
126. Granlund, G.H.: Cognitive vision – background and research issues. Research report, Linköping University (2002)
127. Granrud, C.E., Yonas, A.: Infants' perception of pictorially specified interposition. Journal of Experimental Child Psychology 37, 500–511 (1984)

128. Gray, W.D., Young, R.M., Kirschenbaum, S.S.: Introduction to this special issue on cognitive architectures and human-computer interaction. Human-Computer Interaction 12, 301–309 (1997)
129. Graziano, M.S., Hu, X.T., Gross, C.G.: Coding the locations of objects in the dark. Science 277, 239–241 (1997)
130. Gredebäck, G., von Hofsten, C., Boudreau, P.: Infants' tracking of continuous circular motion and circular motion interrupted by occlusion. Infant Behaviour and Development 144, 1–21 (2002)
131. Grezes, J., Costes, N., Decety, J.: The effects of learning and intention on the neural network involved in the perception of meaningless actions. Brain 122, 1875–1887 (1999)
132. Grush, R.: The emulation theory of representation: motor control, imagery, and perception. Behavioral and Brain Sciences 27, 377–442 (2004)
133. Hadders-Algra, M., Brogren, E., Forssberg, H.: Ontogeny of postural adjustments during sitting in infancy: variation, selection and modulation. Journal of Physiology 493, 273–288 (1996)
134. Haehl, V., Vardaxis, V., Ulrich, B.: Learning to cruise: Bernsteins theory applied to skill acquisition during infancy. Human Movement Science 19, 685–715 (2000)
135. Hainline, L., Riddell, P., Grose-fifer, J., Abramov, I.: Development of accommodation and convergence in infancy. Behavioural Brain Research 49, 30–50 (1992)
136. Haugland, J.: Semantic engines: An introduction to mind design. In: Haugland, J. (ed.) Mind Design: Philosophy, Psychology, Artificial Intelligence, pp. 1–34. Bradford Books, MIT Press, Cambridge, Massachusetts (1982)
137. Hawes, N., Wyatt, J.: Developing intelligent robots with cast. In: Proc. IROS Workshop on Current Software Frameworks in Cognitive Robotics Integrating Different Computational Paradigms (2008)
138. Hawes, N., Wyatt, J., Sloman, A.: An architecture schema for embodied cognitive systems. In: Technical Report CSR-06-12. University of Birmingham, School of Computer Science (2006)
139. Hebb, D.O.: The Organization of Behaviour. John Wiley & Sons, New York (1949)
140. Held, R., Birch, E., Gwiazda, J.: Stereoacuity in human infants. PNAS 77, 5572–5574 (1980)
141. Hermer, L., Spelke, E.S.: Modularity and development: the case of spatial reorientation. Cognition 61, 195–232 (1996)
142. Hersch, M., Billard, A.: Reaching with multi-referential dynamical systems. Autonomous Robots 25, 71–83 (2008)
143. Hespos, S., Gredebäck, G., von Hofsten, C., Spelke, E.S.: Occlusion is hard: Comparing predictive reaching for visible and hidden objects in infants and adults. Cognitive Science 33, 1483–1502 (2009)
144. Hesslow, G.: Conscious thought as simulation of behaviour and perception. Trends in Cognitive Sciences 6(6), 242–247 (2002)
145. von Hofsten, C.: Binocular convergence as a determinant of reaching behaviour in infancy. Perception 6, 139–144 (1977)
146. von Hofsten, C.: Development of visually guided reaching: the approach phase. Journal of Human Movement Studies 5, 160–178 (1979)
147. von Hofsten, C.: Predictive reaching for moving objects by human infants. Journal of Experimental Child Psychology 30, 369–382 (1980)
148. von Hofsten, C.: Eye-hand coordination in newborns. Developmental Psychology 18, 450–461 (1982)
149. von Hofsten, C.: Catching skills in infancy. Experimental Psychology: Human Perception and Performance 9, 75–85 (1983)

150. von Hofsten, C.: Developmental changes in the organization of pre-reaching movements. Developmental Psychology 20, 378–388 (1984)
151. von Hofsten, C.: The early development of the manual system. In: Lindblom, B., Zetterström, R. (eds.) Precursors of Early Speech. Macmillan, Basingstoke (1986)
152. von Hofsten, C.: Structuring of early reaching movements: A longitudinal study. Journal of Motor Behavior 23, 280–292 (1991)
153. von Hofsten, C.: Prospective control: A basic aspect of action development. Human Development 36, 253–270 (1993)
154. von Hofsten, C.: On the early development of predictive abilities. In: Dent, C., Zukow-Goldring, P. (eds.) Evolving Explanations of Development: Ecological approaches to Organism-Environmental Systems, pp. 163–194 (1997)
155. von Hofsten, C.: On the development of perception and action. In: Valsiner, J., Connolly, K.J. (eds.) Handbook of Developmental Psychology, pp. 114–140. Sage, London (2003)
156. von Hofsten, C., Dahlström, E., Fredriksson, Y.: 12-month–old infants' perception of attention direction in static video images. Infancy 8(3), 217–231 (2005)
157. von Hofsten, C., Fazel-Zandy, S.: Development of visually guided hand orientation in reaching. Journal of Experimental Child Psychology 38, 208–219 (1984)
158. von Hofsten, C., Feng, Q., Spelke, E.S.: Object representation and predictive action in infancy. Developmental Science 3, 193–205 (2000)
159. von Hofsten, C., Johansson, K.: Planning to reach for a rotating rod: Developmental aspects. Zeitschrift für Entwickelungspsychologie und Pädagogische Psychologie 41, 207–213 (2009)
160. von Hofsten, C., Kellman, P.J., Putaansuu, J.: Young infants' sensitivity to motion parallax. Infant Behaviour and Development 15, 245–264 (1992)
161. von Hofsten, C., Kochukhova, O., Rosander, K.: Predictive occluder tracking in 4-month-old infants (2006) (submitted manuscript)
162. von Hofsten, C., Lindhagen, K.: Observations on the development of reaching for moving objects. Journal of Experimental Child Psychology 28, 158–173 (1979)
163. von Hofsten, C., Rönnqvist, L.: Preparation for grasping an object: A developmental study. Journal of Experimental Psychology: Human Perception and Performance 14, 610–621 (1988)
164. von Hofsten, C., Rosander, K.: The development of gaze control and predictive tracking in young infants. Vision Research 36, 81–96 (1996)
165. von Hofsten, C., Rosander, K.: Development of smooth pursuit tracking in young infants. Vision Research 37, 1799–1810 (1997)
166. von Hofsten, C., Spelke, E.S.: Object perception and object directed reaching in infancy. Journal of Experimental Psychology: General 114, 198–212 (1985)
167. von Hofsten, C., Vishton, P., Spelke, E.S., Feng, Q., Rosander, K.: Predictive action in infancy: tracking and reaching for moving objects. Cognition 67, 255–285 (1999)
168. von Hofsten, C., Woollacott, M.: Postural preparations for reaching in 9-month-old infants (1990) (unpublished data)
169. Hörnstein, J., Lopes, M., Santos-Victor, J., Lacerda, F.: Sound localization for humanoid robots - building audio-motor maps based on the HRTF. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Beijing, China, pp. 1170–1176 (2006)
170. Horswill, I.: Tagged behavior-based systems: Integrating cognition with embodied activity. IEEE Intelligent Systems, 30–38 (2001)
171. Horswill, I.: Cerebus: A higher-order behavior-based system. AI Magazine 23(1), 27(2002)

172. Huston, S.D., Johnson, J.C.E., Syyid, U.: The ACE Programmer's Guide. Addison-Wesley, Reading (2003)
173. Huttenlocher, P.R.: Morphometric study of human cerebral cortex development. Neuropsychologia 28, 517–527 (1990)
174. Huttenlocher, P.R., Dabholkar, A.S.: Regional differences in synaptogenesis in human cerebral cortex. J. Comp. Neurol. 387, 167–178 (1997)
175. Hyén, D.: The broad frequency-band rotary test, vol. Dissertation Number 152. Linkoping University Medical School, Linkoping, Sweden (1983)
176. Hyvarinen, J., Poranen, A.: Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. Brain 97, 673–692 (1974)
177. Itti, L.: Models of Bottom-Up and Top-Down Visual Attention, Ph.D Thesis. California Institute of Technology (2000)
178. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
179. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience 2, 194–203 (2001)
180. Itti, L., Koch, C.: E: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
181. Ivanenko, Y.P., Dominici, N., Cappellini, G., Lacquantini, F.: Kinematics in newly walking toddlers does not depend upon postural stability. Journal of Neurophysiology 94, 754–763 (2005)
182. Ivanenko, Y.P., Dominici, N., Lacquantini, F.: Development of independent walking in toddlers. Exerc. Sport. Sci. Rev. 35(2), 67–73 (2007)
183. Johansson, R.S., et al.: Eye-hand coordination in object manipulation. Journal of Neuroscience 21, 6917–6932 (2001)
184. Johnson, M.H., Morton, J.: Biology and cognitive development: the case of face recognition. Blackwell, Oxford (1991)
185. Johnson, S.H.: Thinking ahead: the case for motor imagery in prospective judgements of prehension. Cognition 74, 33–70 (2000)
186. Jones, M., Vernon, D.: Using neural networks to learn hand-eye co-ordination. Neural Computing and Applications 2(1), 2–12 (1994)
187. Jonsson, B., von Hofsten, C.: Infants' ability to track and reach for temporarily occluded objects. Developmental Science 6(1), 86–99 (2003)
188. Juscyck, P.W.: Developing phonological categories from the speech signal. In: Ferguson, C.A., Menn, L., Stoel-Gammon, C. (eds.) Phonological Development: Models, Research, Implications, pp. 17–64. York Press, Timonium (1992)
189. Karmiloff-Smith, A.: Beyond Modularity: A developmental perspective on cognitive science. MIT Press, Cambridge (1992)
190. Karmiloff-Smith, A.: Precis of beyond modularity: A developmental perspective on cognitive science. Behavioral and Brain Sciences 17(4), 693–745 (1994)
191. Kayed, N.S., van der Meer, A.: Timing strategies used in defensive blinking to optical collisions in 5- to 7-month-old infants. Infant Behaviour and Development 23, 253–270 (2000)
192. Kellman, P.J., Arterberry, M.E.: The cradle of knowledge. MIT Press, Cambridge (1998)
193. Kellman, P.J., von Hofsten, C., van der Walle, G.A., Condry, K.: Perception of motion and stability during observer motion by pre-stereoscopic infants. In: ICIS, Montreal (1990)
194. Kellman, P.J., Spelke, E.S.: Perception of partly occluded objects in infancy. Cognitive Psychology 15, 438–524 (1983)

195. Kelso, J.A.S.: Dynamic Patterns – The Self-Organization of Brain and Behaviour, 3rd edn. MIT Press, Cambridge (1995)
196. Kieras, D., Meyer, D.: An overview of the epic architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction 12(4) (1997)
197. Kihlstrom, J.F.: The cognitive unconscious. Science 237, 1445–1452 (1987)
198. Kita, S. (ed.): Pointing: Where language, culture and cognition meet. Erlbaum, Mahwah (2003)
199. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4(4), 219–227 (1985)
200. Kochukhova, O., Gredebäck, G.: There are many ways to solve an occlusion task: the role of inertia and recent experience (2006) (submitted manuscript)
201. Kraft, D., Başeski, E., Popović, M., Batog, A.M., Kjr-Nielsen, A., Krüger, N., Petrick, R., Geib, C., Pugeault, N., Steedman, M., Asfour, T., Dillmann, R., Kalkan, S., Wörgötter, F., Hommel, B., Detry, R., Piater, J.: Exploration and planning in a three-level cognitive architecture. In: Proceedings of the First International Conference on Cognitive Systems, Karlsruhe, Germany (2008)
202. Kremenitzer, J.P., Vaughan, H.G., Kurtzberg, D., Dowling, K.: Smooth-pursuit eye movements in the newborn infant. Child Development 50, 442–448 (1979)
203. Krichmar, J.L., Edelman, G.M.: Brain-based devices for the study of nervous systems and the development of intelligent machines. Artificial Life 11, 63–77 (2005)
204. Krichmar, J.L., Edelman, G.M.: Principles underlying the construction of brain-based devices. In: Kovacs, T., Marshall, J.A.R. (eds.) Proceedings of AISB 2006 - Adaptation in Artificial and Biological Systems, Symposium on Grand Challenge 5: Architecture of Brain and Mind, vol. 2, pp. 37–42. University of Bristol, Bristol (2006)
205. Krichmar, J.L., Nitz, D.A., Gally, J.A., Edelman, G.M.: 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)
206. Krichmar, J.L., Reeke, G.N.: The Darwin brain-based automata: Synthetic neural models and real-world devices. In: Reeke, G.N., Poznanski, R.R., Lindsay, K.A., Rosenberg, J.R., Sporns, O. (eds.) Modelling in the neurosciences: from biological systems to neuromimetic robotics, pp. 613–638. Taylor and Francis, Boca Raton (2005)
207. Krichmar, J.L., Seth, A.K., Nitz, D.A., Fleisher, J.G., Edelman, G.M.: Spatial navigation and causal analysis in a brain-based device modelling cortical-hippocampal interactions. Neuroinformatics 3, 197–221 (2005)
208. Kuhl, P.K.: Learning and representation in speech and language. Current opinion in Neurobiology 4, 812–822 (1994)
209. Kuypers, H.G.J.M.: The anatomical organization of the descending pathways and their contribution to motor control especially in primates. In: Desmedt, J.E. (ed.) New Developments in Electromyography and Clinical Neurophysiology, vol. 3, pp. 38–68. S. Karger, New York (1973)
210. Grafton, S.T., Fadiga, L., Arbib, M.A., Rizzolatti, G.: Premotor cortex activation during observation and naming of familiar tools. NeuroImage 6, 231–236 (1997)
211. Laird, J.E., Newell, A., Rosenbloom, P.S.: Soar: an architecture for general intelligence. Artificial Intelligence 33, 1–64 (1987)
212. Landau, B., et al.: Spatial knowledge in a young blind child. Cognition 16, 225–260 (1984)
213. Langley, P.: An cognitive architectures and the construction of intelligent agents. In: Proceedings of the AAAI 2004 Workshop on Intelligent Agent Architectures, Stanford, CA, p. 82 (2004)

214. Langley, P.: An adaptive architecture for physical agents. In: IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pp. 18–25. IEEE Computer Society Press, Compiegne (2005)
215. Langley, P.: Cognitive architectures and general intelligent systems. AI Magazine 27(2), 33–44 (2006)
216. Langley, P., Choi, D., Rogers, S.: Acquisition of hierarchical reactive skills in a unified cognitive architecture. Cognitive Systems Research 10(4), 316–332 (2009)
217. Langley, P., Laird, J.E., Rogers, S.: Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10(2), 141–160 (2009)
218. Lawrence, D.G., Hopkins, D.A.: The development of the motor control in the rhesus monkey: Evidence concerning the role of corticomotoneuronal connections. Brain 99, 235–254 (1976)
219. Ledebt, A., Wiener-Vacher, S.: Head coordination in the sagittal plane in toddlers during walking: preliminary results. Brain Research Bulletin 5/6, 371–373 (1996)
220. Lehman, J.F., Laird, J.E., Rosenbloom, P.S.: A gentle introduction to soar, an architecture for human cognition. In: Sternberg, S., Scarborough, D. (eds.) Invitation to Cognitive Science. Methods, Models, and Conceptual Issues, vol. 4. MIT Press, Cambridge (1998)
221. Leslie, A.M.: Tomm, toby, and agency: Core architecture and domain specificity. In: Hirschfeld, L.A., Gelman, S.A. (eds.) Mapping the Mind: Specificity in Cognition and Culture, pp. 119–148. Cambridge University Press, Cambridge (1994)
222. Lewis, R.L.: Cognitive theory, soar. In: International Encyclopedia of the Social and Behavioural Sciences. Pergamon, Elsevier Science, Amsterdam (2001)
223. Lizowski, U., Carpenter, M., Striano, T., Tomasello, M.: Twelve and 18 month-olds point to provide information. Journal of Cognition and Development 7(2) (2006)
224. Lockman, J.J.: A perception-action perspective on tool use development. Child Development 71, 137–144 (2000)
225. Lockman, J.J., Ashmead, D.H., Bushnell, E.W.: The development of anticipatory hand orientation during infancy. Journal of Experimental Child Psychology 37, 176–186 (1984)
226. Lopes, M., Bernardino, A., Santos-Victor, J., von Hofsten, C., Rosander, K.: Biomimetic eye-neck coordination. In: IEEE International Conference on Development and Learning, Shanghai, China (2009)
227. MacNeilage, P.F., Davis, B.L.: Motor explanations of babbling and early speech patterns. In: Boysson-Bardies, et al. (eds.) Developmental Neurocognition: Speech and Face Processing in the First Year of Life, pp. 341–352. Kluwer Academic Publishers, Amsterdam (1993)
228. Maggiali, M., Cannata, G., Maiolino, P., Metta, G., Randazzo, M., Sandini, G.: Embedded distributed capacitive tactile sensor. In: The 11th Mechatronics Forum Biennial International Conference. University of Limerick, Ireland (2008)
229. von der Malsburg, C., Singer, W.: Principles of cortical network organisations. In: Rakic, P., Singer, W. (eds.) Neurobiology of the Neocortex, pp. 69–99. John Wiley & Sons Ltd., London (1988)
230. Marr, D.: Artificial intelligence – A personal view. Artificial Intelligence 9, 37–48 (1977)
231. Matasaka, N.: From index finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In: Kita, S. (ed.) Pointing: Where language, culture and cognition meet, pp. 69–84. Erlbaum, Mahwah (2003)
232. Matelli, M., Camarda, R., Glickstein, M., Rizzolatti, G.: Afferent and efferent projections of the inferior area 6 in the macaque monkey. J. Comp. Neurol. 251, 281–298 (1986)

233. Matelli, M., Luppino, G., Rizzolatti, G.: Patterns of cytochrome oxidase activity in the frontal agranular cortex of macaque monkey. Behav. Brain Res. 18, 125–137 (1985)
234. Matelli, M., Luppino, G., Rizzolatti, G.: Architecture of superior and mesial area 6 and of the adjacent cingulate cortex. J. Comp. Neurol. 311, 445–462 (1991)
235. Maturana, H.: Biology of cognition. Research Report BCL 9.0, University of Illinois, Urbana, Illinois (1970)
236. Maturana, H.: The organization of the living: a theory of the living organization. Int. Journal of Man-Machine Studies 7(3), 313–332 (1975)
237. Maturana, H., Varela, F.: The Tree of Knowledge – The Biological Roots of Human Understanding. New Science Library, Boston & London (1987)
238. Maturana, H.R., Varela, F.J.: Autopoiesis and Cognition — The Realization of the Living. Boston Studies on the Philosophy of Science. D. Reidel Publishing Company, Dordrecht (1980)
239. Maurer, D.: Infants' perception of faceness. In: Field, T.N., Fox, N. (eds.) Social Perception in Infants, pp. 37–66. Lawrence Erlbaum Associates, Hillsdale (1985)
240. Maurer, D., Lewis, T.L.: Visual acuity: the role of visual input in inducing postnatal change. Clinical Neuroscience Research 1, 239–247 (2001)
241. McCarty, M.E., Clifton, R.K., Ashmead, D.H., Lee, P., Goubet, N.: How infants use vision for grasping objects. Child Development 72(4), 973–987 (2001)
242. McClelland, J.L., NcNaughton, B.L., O'Reilly, R.C.: 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)
243. van der Meer, A.H.L., et al.: Lifting weights in neonates: developing visual control of reaching. Scandinavian Journal of Psychology 37, 424–436 (1996)
244. van der Meer, A.L.H., van der Weel, F., Lee, D.N.: Prospective control in catching by infants. Perception 23, 287–302 (1994)
245. van der Meer, A.L.H., van der Weel, F.R., Lee, D.N.: The functional significance of arm movements in neonates. Science 267, 693–695 (1995)
246. Meltzoff, A.N., Moore, M.K.: Imitation of facial and manual gestures by human neonates. Science 198, 75–78 (1977)
247. Menn, L.: Development of articulatory, phonetic and phonological capabilities. In: Butterworth, B. (ed.) Language Production and Control, vol. 2, pp. 3–50. Academic Press, London (1983)
248. Metcalfe, J.S., McDowell, K., Chang, T.Y., Chen, L.C., Jeka, J.J., Clark, J.E.: 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)
249. Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C., Santos-Victor, J., Bernardino, A., Montesano, L.: The iCub Humanoid Robot: An Open-Systems Platform for Research in Cognitive Development. Submitted to Neural Networks (2010)
250. Metta, G., Sandini, G., Konczak, J.: A developmental approach to visually-guided reaching in artificial systems. Neural Networks 12(10), 1413–1427 (1999)
251. Metta, G., Vernon, D., Sandini, G.: The robotcub approach to the development of cognition: Implications of emergent systems for a common research agenda in epigenetic robotics. In: Proceedings of the Fifth International Workshop on Epigenetic Robotics, EpiRob 2005 (2005)
252. Michel, O.: Webots: professional mobile robot simulation. International Journal of Advanced Robotics Systems 1(1), 39–42 (2004)

253. Milner, A.D., Goodale, M.A.: The Visual Brain in Action. Oxford University Press, Oxford (1995)
254. Minsky, M.: Society of Mind. Simon and Schuster, New York (1986)
255. Montesano, L., Lopes, M.: Learning grasping affordances from local visual descriptors. In: IEEE International Conference on Development and Learning, Shanghai, China (2009)
256. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Modeling affordances using bayesian networks. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, USA (2007)
257. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning object affordances: From sensory motor maps to imitation. IEEE Transactions on Robotics 24(1) (2008)
258. Montesano, L., Lopes, M., Melo, F., Bernardino, A., Santos-Victor, J.: A Computational Model of Object Affordances. IET (2009)
259. Moore, C., Angelopoulos, M., Bennett, P.: The role of movement in the development of joint visual attention. Infant Behavior and Development 20, 83–92 (1997)
260. Morales, M., Mundy, P., Delgado, C.E.F., Yale, M., Messinger, D., Neal, R.: 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)
261. Morales, M., Mundy, P., Delgado, C.E.F., Yale, M., Neal, R., Schwartz, H.K.: Gaze following, temperament, and language development in 6-month-olds: A replication and extension. Infant Behavior and Development 23, 231–236 (2000)
262. Morales, M., Mundy, P., Rojas, J.: Following the direction of gaze and language development in 6-month-olds. Infant Behavior and Development 21, 373–377 (1998)
263. Morissette, P., Ricard, M., D'carie, T.G.: Joint visual attention and pointing in infancy: A longitudinal study of comprehension. British Journal of Developmental Psychology 13, 163–175 (1995)
264. Morse, A., Lowe, R., Ziemke, T.: Towards an enactive cognitive architecture. In: Proceedings of the First International Conference on Cognitive Systems, Karlsruhe, Germany (2008)
265. Mountcastle, V.B., Lynch, J.C.G.A., Georgopoulos, A., Sakata, H., Acuna, C.: Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. Journal of Neurophysiology 38, 871–908 (1975)
266. Müller, M., Wehner, R.: Path integration in desert ants, *cataglyphis fortis*. PNAS 85, 5287–5290 (1988)
267. Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., Rizzolatti, G.: Object representation in the ventral premotor cortex (area f5) of the monkey. J. Neurophysiol. 78, 2226–2230 (1997)
268. Nadel, J., Guerini, C., Peze, A., Rivet, C.: The evolving nature of imitation as a format for communication. In: Nadel, J., Butterworth, G. (eds.) Imitation in Infancy, pp. 209–234. Cambridge University Press, Cambridge (1999)
269. Nanez, J.: Perception of impending collision in 3- to 6-week-old infants. Infant Behaviour and Development 11, 447–463 (1988)
270. Newell, A.: The knowledge level. Artificial Intelligence 18(1), 87–127 (1982)
271. Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge (1990)
272. Newell, A., Simon, H.A.: Computer science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery 19, 113–126 (1976); Tenth Turing award lecture, ACM (1975)

273. Newell, K.M., Scully, D.M., McDonald, P.V., Baillargeon, R.: Task constraints and infant grip configurations. Developmental Psychobiology 22, 817–832 (1989)
274. Nilsson, L.: A Child is Born. Albert Bonniers förlag
275. Ogden, B., Dautenhahn, K., Stribling, P.: Interactional structure applied to the identification and generation of visual interactive behaviour: Robots that (usually) follow the rules. In: Wachsmuth, I., Sowa, T. (eds.) GW 2001. LNCS (LNAI), vol. 2298, pp. 254–268. Springer, Heidelberg (2002)
276. Okano, K., Tanji, J.: Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. Exp. Brain Res. 66, 155–166 (1987)
277. Okuma, K., Taleghani, A., de Freitas, N., Little, J., Lowe, D.: A boosted particle filter: Multitarget detection and tracking. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 28–39. Springer, Heidelberg (2004)
278. Olsson, L., Nehaniv, C.L., Polani, D.: From unknown sensors and actuators to actions grounded in sensorimotor perceptions. Connection Science 18(2) (2006)
279. Örnkloo, H., von Hofsten, C.: Fitting objects into holes: on the development of spatial cognition skills. Developmental Psychology 43, 403–416 (2007)
280. Paolo, E.D., Rohde, M., Jaegher, H.D.: Horizons for the enactive mind: Values, social interaction, and play. In: Stewart, J., Gapenne, O., Paolo, E.D. (eds.) Enaction: Towards a New Paradigm for Cognitive Science. MIT Press, Cambridge (2008)
281. Parsons, L.M., Fox, P.T., Downs, J.H., Glass, T., Hirsch, T.B., Martin, C.C., Jerabek, P.A., Lancaster, J.L.: Use of implicit motor imagery for visual shape discrimination as revealed by PET. Nature 375, 54–58 (1995)
282. Pavel, M.: Predictive control of eye movement. In: Kowler, E. (ed.) Eye Movements and Their Role in Visual and Cognitive Processes, Reviews of Oculomotor Research, vol. 4, pp. 71–114. Elsevier, Amsterdam (1990)
283. Pellegrino, G.D., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G.: Understanding motor events: a neurophysiological study. Exp. Brain Res. 91, 176–180 (1992)
284. Penfield, W., Rasmussen, T.: The Cerebral Cortex of Man. A Clinical Study of Localization of function. Macmillan, New York (1950)
285. Petrick, R.P.A., Bacchus, F.: A knowledge-based approach to planning with incomplete information and sensing. In: Proceedings of the International Conference on Artificial Intelligence Planning and Scheduling, pp. 212–221. AAAI Press, Menlo Park (2002)
286. Petrick, R.P.A., Bacchus, F.: Extending the knowledge-based approach to planning with incomplete information and sensing. In: Proceedings of the International Conference on Artificial Intelligence Planning and Scheduling, pp. 2–11 (2004)
287. Petrides, M., Pandya, D.N.: Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. J. Comp. Neurol. 228, 105–116 (1984)
288. Piaget, J.: The origins of intelligence in the child. Routledge, New York (1953)
289. Piaget, J.: The construction of reality in the child. Basic Books, New York (1954)
290. Piaget, J.: The Construction of Reality in the Child. Routeledge and Kegan Paul, London (1955)
291. Pinker, S.: Visual cognition: An introduction. Cognition 18, 1–63 (1984)
292. Pinker, S.: How the Mind Works. W. W. Norton and Company, New York (1997)
293. Pivonelli, D.J., Bering, J.M., Giambrone, S.: Chimpanzees ”pointing”: Another error of the argument by analogy? In: Kita, S. (ed.) Pointing: Where language, culture and cognition meet, pp. 33–68. Erlbaum, Mahwah (2003)
294. Posner, M.I., Dehaene, S.: Attentional networks. Trends Neurosci. 17, 75–79 (1994)
295. Posner, M.I., Petersen, S.E.: The attention system of the human brain. Annu. Rev. Neurosci. 13, 25–42 (1990)

296. Posner, M.I., Petersen, S.E., Fox, P.T., Raichle, M.E.: Localization of cognitive operations in the human brain. Science 240, 1627–1631 (1988)
297. Posner, M.I., Rafal, R.D., Choate, L.S., Vaughan, J.: Inhibition of return: Neural basis and function. Cognitive Neuropsychology 2, 211–228 (1985)
298. Pylyshyn, Z.W.: Computation and Cognition, 2nd edn. Bradford Books, MIT Press, Cambridge, MA (1984)
299. Ramamurthy, U., Baars, B., D'Mello, S.K., Franklin, S.: LIDA: A working model of cognition. In: Fum, D., Missier, F.D., Stocco, A. (eds.) Proceedings of the 7th International Conference on Cognitive Modeling, pp. 244–249 (2006)
300. Ramsay, D.S.: Fluctuations in unimanual hand preference in infants following the onset of duplicated syllable babbling. Developmental Psychology 21, 318–324 (1985)
301. Reddy, B.S., Chatterji, B.N.: An FFT-based technique for translation, rotation, and scale0-invariant image registration. IEEE Trans. Image Processing 5(8), 1266–1271 (1996)
302. Reed, E.S.: Encountering the world: towards an ecological psychology. Oxford University Press, New York (1996)
303. Regolin, L., et al.: Object and spatial representations in detour problems by chicks. Animal Behaviour 49, 195–199 (1995)
304. Reichardt, W.E.: Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In: Principles of Sensory Communications, New York, pp. 303–317 (1961)
305. Retaux, S., Harris, W.A.: Engrailed and retinotectal topography. Trends in Neuroscience 19, 542–546 (1996)
306. Righetti, L., Ijspeert, A.: Design methodologies for central pattern generators: an application to crawling humanoids. In: Proceedings of Robotics: Science and Systems, Philadelphia, USA, pp. 191–198 (2006)
307. Righetti, L., Ijspeert, A.: Pattern generators with sensory feedback for the control of quadruped locomotion. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation, ICRA 2008 (2008)
308. Righetti, L., Nylen, A., Rosander, K., Ijspeert, A.: Is the locomotion of crawling infants different from other quadrupeded mammals? (submitted, 2010)
309. Ritter, F.E., Young, R.M.: Introduction to this special issue on using cognitive models to improve interface design. International Journal of Human-Computer Studies 55, 1–14 (2001)
310. Rizzolatti, G., Camarda, R.: Neural circuits for spatial attention and unilateral neglect. In: Jeannerod, M. (ed.) Neurophysiological and neuropsychological aspects of spatial neglect, pp. 289–313. North-Holland, Amsterdam (1987)
311. Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., Matelli, M.: Functional organization of inferior area 6 in the macaque monkey: Ii. area f5 and the control of distal movements. Exp. Brain Res. 71, 491–507 (1988)
312. Rizzolatti, G., Craighero, L.: The mirror neuron system. Annual Review of Physiology 27, 169–192 (2004)
313. Rizzolatti, G., Fadiga, L.: Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area f5). In: Bock, G.R., Goode, J.A. (eds.) Sensory Guidance of Movement, Novartis Foundation Symposium 218, pp. 81–103. John Wiley and Sons, Chichester (1998)
314. Rizzolatti, G., Fadiga, L., Fogassi, L., Gallese, V.: The space around us. Science, 190–191 (1997)
315. Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L.: Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3, 131–141 (1996)

316. Rizzolatti, G., Fogassi, L., Gallese, V.: Parietal cortex: from sight to action. Current Opinion in Neurobiology 7, 562–567 (1997)
317. Rizzolatti, G., Gentilucci, M.: Motor and visual-motor functions of the premotor cortex. In: Rakic, P., Singer, W. (eds.) Neurobiology of Neocortex, pp. 269–284. John Wiley and Sons, Chichester (1988)
318. Rizzolatti, G., Riggio, L., Dascola, I., Umilta, C.: Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia 25, 31–40 (1987)
319. Rizzolatti, G., Riggio, L., Sheliga, B.M.: Space and selective attention. In: Umiltà, C., Moscovitch, M. (eds.) Attention and performance XV, pp. 231–265. MIT Press, Cambridge (1994)
320. Robin, D.J., Berthier, N.E., Clifton, R.K.: Infants' predictive reaching for moving objects in the dark. Developmental Psychology 32, 824–835 (1996)
321. Rochat, P.: 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)
322. Rochat, P., Goubet, N.: Development of sitting and reaching in 5- to 6-month-old infants. Infant Behavior and Development 18, 53–68 (1995)
323. Rosander, K., Gredebäck, G., Nystroöm, P., von Hofsten, C.: (2006) (submitted manuscript)
324. Rosander, K., von Hofsten, C.: Visual-vestibular interaction in early infancy. Exp. Brain Res. 133, 321–333 (2000)
325. Rosander, K., von Hofsten, C.: Infants' emerging ability to represent object motion. Cognition 91, 1–22 (2004)
326. Rosenbloom, P., Laird, J., Newell, A. (eds.): The Soar Papers: Research on Integrated Intelligence. MIT Press, Cambridge (1993)
327. Rozin, P.: The evolution of intelligence and access to cognitive unconscious. Psychobiology and Physiological Psychology 6, 245–279 (1976)
328. Ruesch, J., Lopes, M., Hornstein, J., Santos-Victor, J., Pfeifer, R.: Multimodal saliency-based bottom-up attention - a framework for the humanoid robot icub. In: Proc. International Conference on Robotics and Automation, Pasadena, CA, USA, pp. 962–967 (2008)
329. Sandini, G., Metta, G., Vernon, D.: Robotcub: An open framework for research in embodied cognition. In: IEEE-RAS/RSJ International Conference on Humanoid Robots (Humanoids 2004), pp. 13–32 (2004)
330. Sandini, G., Metta, G., Vernon, D.: The icub cognitive humanoid robot: An open-system research platform for enactive cognition. In: Lungarella, M., Iida, F., Bongard, J.C., Pfeifer, R. (eds.) 50 Years of Aritficial Intelligence. LNCS (LNAI), vol. 4850, pp. 359–370. Springer, Heidelberg (2007)
331. Sanefuji, W., Ohgami, H., Hashiya, K.: Detection of the relevant type of locomotion in infancy: crawlers versus walkers. Infant Behavior and Development 31(4), 624–628 (2008)
332. Scaife, M., Bruner, J.S.: The capacity for joint visual attention in infants. Nature 53, 265–266 (1975)
333. Scassellati, B.: Theory of mind for a humanoid robot. Autonomous Robots 12, 13–24 (2002)
334. Seth, A., McKinstry, J., Edelman, G., Krichmar, J.L.: Active sensing of visual and tactile stimuli by brain-based devices. International Journal of Robotics and Automation 19(4), 222–238 (2004)

335. Shanahan, M.P.: Cognition, action selection, and inner rehearsal. In: Proceedings IJCAI Workshop on Modelling Natural Action Selection, pp. 92–99 (2005)
336. Shanahan, M.P.: Emotion, and imagination: A brain-inspired architecture for cognitive robotics. In: Proceedings AISB 2005 Symposium on Next Generation Approaches to Machine Consciousness, pp. 26–35 (2005)
337. Shanahan, M.P.: A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition 15, 433–449 (2006)
338. Shanahan, M.P., Baars, B.: Applying global workspace theory to the frame problem. Cognition 98(2), 157–176 (2005)
339. Shapiro, S.C., Bona, J.P.: The GLAIR cognitive architecture. In: Samsonovich, A. (ed.) Biologically Inspired Cognitive Architectures-II: Papers from the AAAI Fall Symposium, Technical Report FS-09-01, pp. 141–152. AAAI Press, Menlo Park (2009)
340. Shatz, C.J.: The developing brain. Scientific American, 35–41 (1992)
341. Sheliga, B.M., Riggio, L., Craighero, L., Rizzolatti, G.: Spatial attention and eye movements. Exp. Brain Res. 105, 261–275 (1995)
342. Sheliga, B.M., Riggio, L., Craighero, L., Rizzolatti, G.: Spatial attention-determined modifications in saccade trajectories. Neuroreport 6, 585–588 (1995)
343. Shepard, R.N., Hurwitz, S.: Upward direction, mental rotation, and discrimination of left and right turns in maps. Cognition 18, 161–193 (1984)
344. Siddiqui, A.: Object size as a determinant of grasping in infancy. Journal of Genetic Psychology 156, 345–358 (1995)
345. Simion, F., Regolin, L., Bulf, H.: A predispostion for biological motion in the newborn infant. PNAS 105, 809–813 (2008)
346. Sirigu, A., Duhamel, J.R., Cohen, L., Pillon, B., Dubois, B., Agid, Y.: The mental representation of hand movements after parietal cortex damage. Science 273, 1564–1568 (1996)
347. Sloman, A.: Varieties of affect and the cogaff architecture schema. In: Proceedings of the AISB 2001 Symposium on Emotion, Cognition, and Affective Computing, York, UK (2001)
348. Sloman, A.: What's a Research Roadmap for? Why do we need one? How can we produce one? (2007), `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0701` (Online; accessed August 10, 2010)
349. Smith, W.C., Johnson, S.P., Spelke, E.S.: Motion and edge sensitivity in perception of object unity. Cognitive Psychology 46, 31–64 (2002)
350. Speidel, G.S.: Imitation: a bootstrap for learning to speak. In: Speidel, G.E., Nelson, K.E. (eds.) The many faces of imitation in language learning, pp. 151–180. Springer, Heidelberg (1989)
351. Spelke, E.S.: Principles of object perception. Cognitive Science 14, 29–56 (1990)
352. Spelke, E.S.: Core knowledge. American Psychologist, 1233–1243 (2000)
353. Spelke, E.S.: Core knowledge. In: Kanwisher, N., Duncan, J. (eds.) Attention and Performance, vol. 20. Oxford University Press, Oxford (2003)
354. Spelke, E.S., von Hofsten, C.: Predictive reaching for occluded objects by six-month-old infants. Journal of Cognition and Development 2, 261–282 (2001)
355. Spelke, E.S., von Hofsten, C., Kestenbaum, R.: Object perception and object-directed reaching in infancy: interaction of spatial and kinetic information for object boundaries. Developmental Psychology 25, 185–196 (1989)
356. Spelke, E.S., de Walle, G.V.: Perceiving and reasoning about objects: insights from infants. In: Eilan, N., McCarthy, R., Brewer, W. (eds.) Spatial Representation. Blackwell, Oxford (1993)

357. de Sperati, C., Stucchi, D.: Recognizing the motion of a graspable object is guided by handedness. Neuroreport 8, 2761–2765 (1997)
358. Starkey, P., Spelke, E.S., Gelman, R.: Numerical abstraction by human infants. Cognition 36, 97–127 (1990)
359. Stewart, J., Gapenne, O., Paolo, E.A.D.: Enaction: Toward a New Paradigm for Cognitive Science. MIT Press, Cambridge (2011)
360. Striano, T., Rochat, P.: Socio-emotional development in the first year of life. In: Rochat, P. (ed.) Early social cognition. Erlbaum, Mahwah (1999)
361. Striano, T., Tomasello, M.: Infant development: Physical and social cognition. In: International encyclopedia of the social and behavioral sciences, pp. 7410–7414 (2001)
362. Sun, R.: A tutorial on CLARION. In: Cognitive Science Department. Rensselaer Polytechnic Institute (2003), `http://www.cogsci.rpi.edu/rsun/sun.tutorial.pdf`
363. Sun, R.: Desiderata for cognitive architectures. Philosophical Psychology 17(3), 341–373 (2004)
364. Sun, R.: The importance of cognitive architectures: an analysis based on clarion. Journal of Experimental & Theoretical Artificial Intelligence 19(2), 159–193 (2007)
365. Swain, M., Ballard, D.: Color indexing. Internation Journal of Computer Vision 7(1), 11–32 (1991)
366. Swain, M.J., Ballard, D.H.: Indexing via colour histograms. In: International Conference on Computer Vision – ICCV 1990, pp. 390–393 (1990)
367. Thelen, E.: Time-scale dynamics and the development of embodied cognition. In: Port, R.F., van Gelder, T. (eds.) Mind as Motion – Explorations in the Dynamics of Cognition, pp. 69–100. Bradford Books, MIT Press, Cambridge, Massachusetts (1995)
368. Thelen, E., Corbetta, D., Spencer, J.P.: Development of reaching during the first year: Role of movement speed. Journal of Experimental Psychology: Human Perception and Performance 22, 1059–1076 (1996)
369. Thelen, E., Fischer, D.M., Ridley-Johnson, R.: The relationship between physical growth and a newborn reflex. Infant Behavior and Development 7, 479–493 (1984)
370. Thelen, E., Smith, L.B.: A Dynamic Systems Approach to the Development of Cognition and Action. MIT Press / Bradford Books Series in Cognitive Psychology. MIT Press, Cambridge (1994)
371. Thelen, E., Smith, L.B.: Development as a dynamic system. Trends Cognitive Science 7, 343–348 (2003)
372. Thelen, E., Corbetta, D., Kamm, K., Spencer, I.P., Schneider, K., Zernicker, R.F.: The transition to reaching: Mapping intention and intrinsic dynamics. Child Development 64, 1058–1099 (1993)
373. Thompson, E.: Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Harvard University Press, Boston (2007)
374. Tipper, S.P., Lortie, C., Baylis, G.C.: Selective reaching: evidence for action-centered attention. J. Exp. Psychol. Hum. Percept. Perform 18, 891–905 (1992)
375. Tomasello, M.: Acquiring linguistic constructions. In: Kuhn, D., Siegler, R. (eds.) Handbook of Child Psychology. Wiley, New York (2006)
376. Tomasello, M., Carpenter, M., Liszkowski, U.: A new look at infant pointing. Child Development 78(3), 705–722 (2007)
377. Trevarthen, C., Kokkinaki, T., Fiamenghi Jr., G.A.: What infants' imitations communicate: with mothers, with fathers and with peers. In: Nadel, J., Butterworth, G. (eds.) Imitation in Infancy, pp. 61–124. Cambridge University Press, Cambridge (1999)
378. Umilta, M.A., et al.: I know what you are doing: A neurophysiological study. Neuron 31, 155–165 (2001)

379. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (eds.) Analysis of visual behavior, pp. 549–586. MIT Press, Cambridge (1982)
380. Varela, F.: Principles of Biological Autonomy. Elsevier North Holland, New York (1979)
381. Varela, F., Thompson, E., Rosch, E.: The Embodied Mind. MIT Press, Cambridge (1991)
382. Varela, F.J.: Whence perceptual meaning? A cartography of current ideas. In: Varela, F.J., Dupuy, J.P. (eds.) Understanding Origins – Contemporary Views on the Origin of Life, Mind and Society, Boston Studies in the Philosophy of Science, pp. 235–263. Kluwer Academic Publishers, Dordrecht (1992)
383. Vernon, D.: The space of cognitive vision. In: Christensen, H.I., Nagel, H.H. (eds.) Cognitive Vision Systems. LNCS, vol. 3948, pp. 7–26. Springer, Heidelberg (2006)
384. Vernon, D.: Cognitive vision: The case for embodied perception. Image and Vision Computing 26(1), 127–141 (2008)
385. Vernon, D.: Enaction as a conceptual framework for development in cognitive robotics. Paladyn. Journal of Behavioral Robotics 1(2), 89–98 (2010)
386. Vernon, D., Furlong, D.: Philosophical foundations of enactive AI. In: Lungarella, M., Iida, F., Bongard, J.C., Pfeifer, R. (eds.) 50 Years of Aritficial Intelligence. LNCS (LNAI), vol. 4850, pp. 53–62. Springer, Heidelberg (2007)
387. Vernon, D., Metta, G., Sandini, G.: A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. IEEE Transaction on Evolutionary Computation 11(2), 151–180 (2007)
388. Vogt, O., Vogt, C.: Ergebnisse unserer hirnforschung. J. Psychol. Neurol. 25, 277–462 (1919)
389. de Vries, J.I.P., Visser, G.H.A., Prechtl, H.F.R.: The emergence of fetal behaviour. i. qualitative aspects. Early Human Development 23, 159–191 (1982)
390. Vygotsky, L.: Mind in society: The development of higher psychological processes. Harvard University Press, Cambridge (1978)
391. Wang, R.F., Spelke, E.S.: Human spatial representation: insights from animals. Trends in Cognitive Sciences 6, 376–382 (2002)
392. Watkins, C.J.C.H.: Learning from delayed rewards. Doctoral Thesis, Cambridge University, Cambridge, England (1989)
393. Weng, J.: A theory for mentally developing robots. In: Proceedings of the 2nd International Conference on Development and Learning (ICDL 2002). IEEE Computer Society, Los Alamitos (2002)
394. Weng, J.: Developmental robotics: Theory and experiments. International Journal of Humanoid Robotics 1(2), 199–236 (2004)
395. Weng, J.: A theory of developmental architecture. In: Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004), La Jolla (2004)
396. Weng, J., Hwang, W., Zhang, Y., Yang, C., Smith, R.: Developmental humanoids: Humanoids that develop skills automatically. In: Proceedings the First IEEE-RAS International Conference on Humanoid Robots, Cambridge, MA (2000)
397. Weng, J., Zhang, Y.: Developmental robots - a new paradigm. In: Proc. Second International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems (2002)
398. Wilcox, T., Woods, R., Chapa, C., McCurry, S.: Multisensory exploration and object individuation in infancy. Developmental Psychology 43, 479–495 (2007)
399. Wing, A.M., Turton, A., Fraser, C.: Grasp size and accuracy of approach in reaching. Journal of motor Behavior 18, 245–261 (1986)

400. Winograd, T., Flores, F.: Understanding Computers and Cognition – A New Foundation for Design. Addison-Wesley Publishing Company, Inc., Reading (1986)
401. Witherington, D.C.: The development of prospective grasping control between 5 and 7 months: A longitudinal study. Infancy 7(2), 143–161 (2005)
402. Witherington, D.C., et al.: The development of anticipatory postural adjustments in infancy. Infancy 3, 495–517 (2002)
403. Wolff, P.H.: The development of behavioral states and the expression of emotions in early infancy. Chicago University Press, Chicago (1987)
404. Woodward, A.L., Guajardo, J.J.: Infants' understanding of the point gesture as an object-directed action. Cognitive Development 17, 1061–1084 (2002)
405. Woollacott, M., Debu, M., Mowatt, M.: Neuromuscular control of posture in the infant and child: Is vision dominant? Journal of Motor Behavior 19, 167–186 (1987)
406. Woolsey, C.N.: Organization of somatic sensory and motor areas of the cerebral cortex. In: Harlow, H.F., Woolsey, C.N. (eds.) Biological and Biochemical Bases of Behavior, pp. 63–81. University of Wisconsin Press, Madison (1958)
407. Wray, R., Chong, R., Phillips, J., Rogers, S., Walsh, B., Laird, J.E.: A survey of cognitive and agent architectures (2010), `http://ai.eecs.umich.edu/cogarch0/` (online; accessed August 10, 2010)
408. Wynn, K.: Addition and subtraction in infants. Nature 358, 749–750 (1992)
409. Xu, F., Spelke, E.S.: Large number discrimination in 6-month-old infants. Cognition 74, B1–B11 (2000)
410. Yonas, A., Arterberry, M.E., Granrud, C.E.: Space perception in infancy. In: Vasta (ed.) Annals of Child Development, pp. 1–34. JAI Press, Greenwich (1987)
411. Yonas, A., Pettersen, L., Lockman, J.J.: Infants' sensitivity to optical information for collision. Canadian Journal of Psychology 33, 268–276 (1979)
412. Yuodelis, C., Hendrickson, A.: A qualitative and quantitative analysis of the human fovea during development. Vision Res. 26, 847–855 (1986)
413. Ziemke, T.: Are robots embodied? In: Balkenius, C., Zlatev, J., Dautenhahn, K., Kozima, H., Breazeal, C. (eds.) Proceedings of the First International Workshop on Epigenetic Robotics — Modeling Cognitive Development in Robotic Systems, Lund, Sweden, Lund University Cognitive Studies, vol. 85, pp. 75–83 (2001)
414. Ziemke, T.: What's that thing called embodiment? In: Alterman, R., Kirsh, D. (eds.) Proceedings of the 25th Annual Conference of the Cognitive Science Society, Lund University Cognitive Studies, pp. 1134–1139. Lawrence Erlbaum, Mahwah (2003)
415. Ziemke, T., Lowe, R.: On the role of emotion in embodied cognitive architectures: From organisms to robots. Cognition and Computation 1, 104–117 (2009)
416. Zoia, S., Blasen, L., Dttavio, G., Bulgheroni, M., Pezzetta, E., Scatar, A., Castielo, U.: Evidence of early of action planning in the human foetus: a kinematic study. EBR 176, 217–226 (2007)

Index

The Cognitive Systems Monographs (COSMOS) publish new developments and advances in the fields of cognitive systems research, rapidly and informally but with a high quality. The intent is to bridge cognitive brain science and biology with engineering disciplines. It covers all the technical contents, applications, and multidisciplinary aspects of cognitive systems, such as Bionics, System Analysis, System Modelling, System Design, Human Motion, Understanding, Human Activity Understanding, Man-Machine Interaction, Smart and Cognitive Environments, Human and Computer Vision, Neuroinformatics, Humanoids, Biologically motivated systems and artefacts Autonomous Systems, Linguistics, Sports Engineering, Computational Intelligence, Biosignal Processing, or Cognitive Materials as well as the mothodologies behind them. Within the scope of the series are monographs, lecture notes, selected contributions from specialized conferences and workshops, as well as selected PhD theses.

»COSMOS

11

Vernon et al.

David Vernon • Claes von Hofsten • Luciano Fadiga

A Roadmap for Cognitive Development in Humanoid Robots

This book addresses the central role played by development in cognition. The focus is on applying our knowledge of development in natural cognitive systems, specifically human infants, to the problem of creating artificial cognitive systems in the guise of humanoid robots. The approach is founded on the three-fold premise that (a) cognition is the process by which an autonomous self-governing agent acts effectively in the world in which it is embedded, (b) the dual purpose of cognition is to increase the agent's repertoire of effective actions and its power to anticipate the need for future actions and their outcomes, and (c) development plays an essential role in the realization of these cognitive capabilities. Our goal in this book is to identify the key design principles for cognitive development. We do this by bringing together insights from four areas: enactive cognitive science, developmental psychology, neurophysiology, and computational modelling. This results in roadmap comprising a set of forty-three guidelines for the design of a cognitive architecture and its deployment in a humanoid robot. The book includes a case study based on the iCub, an open-systems humanoid robot which has been designed specifically as a common platform for research on embodied cognitive systems.

ISBN 978-3-642-16903-8

9 783642 169038

springer.com

A Roadmap for Cognitive Development in Humanoid Robots