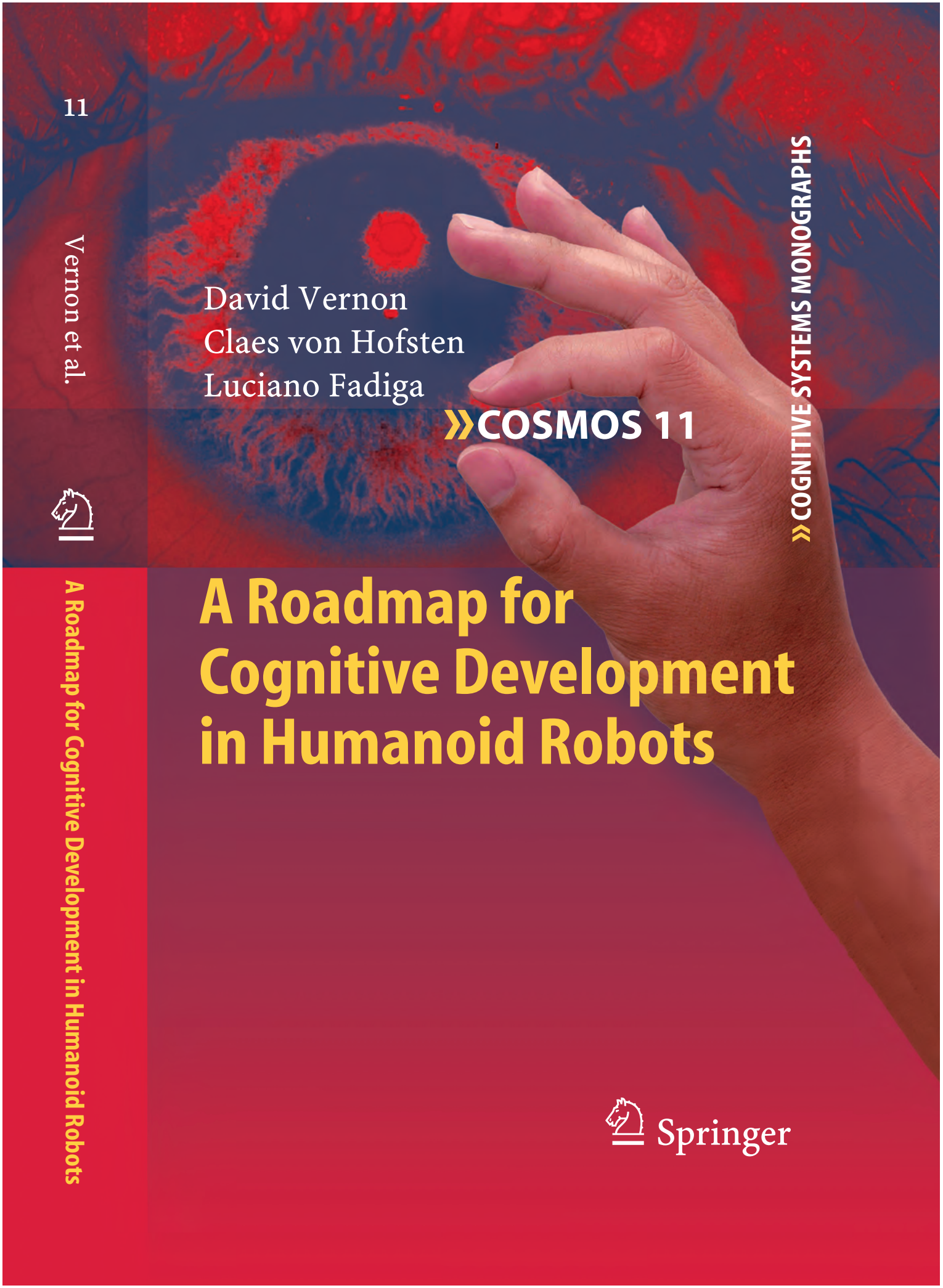

11
Vernon et al.
A Roadmap for Cognitive Development in Humanoid Robots
David Vernon
Claes von Hofsten
Luciano Fadiga
COSMOS 11
COGNITIVE SYSTEMS MONOGRAPHS
A Roadmap for Cognitive Development in Humanoid Robots
Springer

Cognitive Systems Monographs

Volume 11

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

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.

We will now introduce a framework which encapsulates all these considerations.

1.3 Enaction

There are many alternative perspectives on cognition: what it is, why it is necessary, and how it is achieved. We have already argued that cognition arises from an agent's need to compensate for latencies in neural processing by anticipating what may be about to happen and by preparing its actions accordingly. So we can agree fairly easily what cognition is — a process of anticipating events and acting appropriately and effectively — and why it is necessary — to overcome the physical limitations of biological brains and the limitations of bodily movements operating in a dynamic environment. The difficulty arises when we consider how cognition is achieved. There are several competing theories of cognition, each of which makes its own set of assumptions. Here, we wish to focus on one particularly important paradigm — enaction — and pick out its most salient aspects in order to provide a sound theoretical foundation for the role of development in cognition [235, 236, 237, 238, 359, 380, 382, 381, 400] .

The principal idea of enaction is that a cognitive system develops it own understanding of the world around it through its interactions with the environment. Thus, enaction entails that the cognitive system operates autonomously and that it generates its own models of how the world works. When dealing with enactive systems, there are five key elements to consider [280, 373]:

1. Autonomy
2. Embodiment
3. Emergence
4. Experience
5. Sense-making

Autonomy is the self-maintaining organizational characteristic of living creatures that enables them to use their own capacities to manage their interactions with the world, and with themselves, in order to remain viable [61, 108]. This simply means that the system is entirely self-governing and self-regulating: it is not controlled by any outside agency and this allows it to stand apart from the rest of the environment and operate independently of it. That's not to say that the system isn't influenced by the world around it, but rather that these influences are brought about through interactions that don't threaten the autonomous operation of the system.

The second element of enaction is the idea of embodiment. For our purposes here, embodiment means that the system must exist in the world as a physical entity which can interact directly with the environment. This means the system can act on things in the world around it and they, in turn, can act on the system. These things can be inanimate objects or animate agents, cognitive or not. As it happens, there

are some subtle distinctions which can be made about different types of embodiment and we will return to this topic later in the chapter in Section 1.4.

The element of emergence refers to the manner in which cognition arises in the system. Specifically, it refers to the laws and mechanisms which govern the behaviour of the component parts of the system. In an emergent system, the behaviour we call cognition arises from the dynamic interplay between the components and the laws and mechanisms we mentioned govern only the behaviour of the component parts; they don't specify the behaviour of the interplay between the components. Thus, behaviour emerges indirectly because of the internal dynamics. Crucially, these internal dynamics must maintain the autonomy of the system and, as we will see shortly, they also condition the experiences of the system through their embodiment in a specific structure.

Experience is the fourth element of enaction. This is nothing more than the cognitive system's history of interaction with the world around it: the actions it takes in the environment and the actions arising in the environment which impinge on the cognitive system. These interactions don't control the system (otherwise it wouldn't be autonomous) but they do trigger changes in the state of the system. The changes that can be triggered are *structurally determined*: they depend on the system structure, i.e. the embodiment of the self-organizational principles that make the system autonomous. This structure is also referred to as the system's phylogeny: the innate capabilities of an autonomous system with which it is equipped at the outset and which form the basis for its continued existence. The experience of the systems — its history of interactions — involving *structural coupling* between the system and its environment is referred to as its ontogeny.

Finally, we come to the fifth and, arguably, the most important element of enaction: sense-making. This term refers to the relationship between the knowledge encapsulated by a cognitive system and the interactions which gave rise to it. In particular, it refers to the idea that this emergent knowledge is generated by the system itself and that it captures some regularity or lawfulness in the interactions of the system, i.e. its experience. However, the sense it makes is dependent on the way in which it can interact: its own actions and its perceptions of the environment's action on it. Since these perceptions and actions are the result of an emergent dynamic process that is first and foremost concerned with maintaining the autonomy and operational identity of the system, these perceptions and actions are unique to the system itself and the resultant knowledge makes sense only insofar as it contributes to the maintenance of the system's autonomy. This ties in neatly with our view of cognition, the role of which is to anticipate events and increase the space of actions in which a system can engage. By making sense of its experience, the cognitive system is constructing a model that has some predictive value, exactly because it captures some regularity or lawfulness in its interactions. This self-generated model of the system's experience lends the system greater flexibility in how it interacts in the future. In other words, it endows the system with a larger repertoire of possible actions that allow richer interactions, increased perceptual capacity, and the possibility of constructing even better models that encapsulate knowlege with even greater predictive power. And so it goes, in a virtuous circle. Note that this

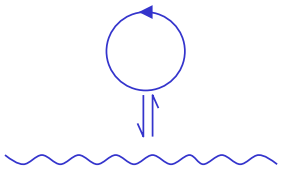


Fig. 1.1 Maturana and Varela's ideograms to denote structurally-determined autopoietic and organizationally-closed systems. The arrow circle denotes the autonomy, self-organization, and self-production of the system, the rippled line the environment, and the bi-directional half-lines the mutual perturbation — structural coupling — between the two.

sense-making and the resultant knowledge says nothing at all about what is really out there in the environment; it doesn't have to. All it has to do is make sense for the continued existence and autonomy of the cognitive system. Sense-making is actually the source of the term enaction. In making sense of its experience, the cognitive system is somehow bringing out through its actions — enacting — what is important for the continued existence of the system. This enaction is effected by the system as it is embedded in its environment, but as an autonomous entity distinct from the environment, through an emergent process of making sense of its experience. This sense-making is, in fact, cognition [108].

The founders of the enactive approach, Maturana and Varela, introduced a diagrammatic way of conveying the self-organizing and self-maintaining autonomous nature of an enactive system, perturbing and being perturbed by its environment [237]: see Fig. 1.1. The arrowed circle denotes the autonomy and self-organization of the system, the rippled line the environment, and the bi-directional half-arrows the mutual perturbation.

1.3.1 Enaction and Development

So what has all this to do with development? Our position in this book is that learning arises as a consequence of the interaction between the cognitive agent and the world around it, whereas development arises from learning through a process of interaction of the cognitive agent with itself. We remarked above that the process of sense-making forms a virtuous circle in that the self-generated model of the system's experience provides a larger repertoire of possible actions, richer interactions, increased perceptual capacity, and potentially better self-generated models, and so on. Recall also our earlier remarks that the cognitive system's knowledge is represented by the state of the system. When this state is embodied in the system's central nervous system, the system has much greater plasticity in two senses: (a) the nervous system can accommodate a much larger space of possible associations between system-environment interactions, and (b) it can accommodate a much larger space of potential actions. Consequently, the process of cognition involves the system modifying its own state, specifically its central nervous system, as it enhances

its predictive capacity and its action capabilities. This is exactly what we mean by development. This generative (i.e. self-constructed) autonomous learning and development is one of the hallmarks of the enactive approach [280, 108].

Development, then, is identically the cognitive process of establishing and enlarging the possible space of mutually-consistent couplings in which a system can engage (or, perhaps more appropriately, which it can withstand without compromising its autonomy). The space of perceptual possibilities is predicated not on an absolute objective environment, but on the space of possible actions that the system can engage in whilst still maintaining the consistency of the coupling with the environment. These environmental perturbations don't control the system since they are not components of the system (and, by definition, don't play a part in the self-organization) but they do play a part in the ontogenetic development of the system. Through this ontogenetic development, the cognitive system develops its own epistemology, i.e. its own system-specific history- and context-dependent knowledge of its world, knowledge that has meaning exactly because it captures the consistency and invariance that emerges from the dynamic self-organization in the face of environmental coupling. Again, it comes down to the preservation of autonomy, but this time doing so in an every increasing space of autonomy-preserving couplings.

This process of development is achieved through self-modification by virtue of the presence of a central nervous system: not only does environment perturb the system (and vice versa) but the system also perturbs itself and the central nervous system adapts as a result. Consequently, the system can develop to accommodate a much larger space of effective system action. This is captured in a second ideogram of Maturana and Varela (see Fig. 1.2) which adds a second arrow circle to the ideogram to depict the process of development through self-perturbation and self-modification. In essence, development *is* autonomous self-modification and requires the existence of a viable phylogeny, including a nervous system, and a suitable ontogeny.

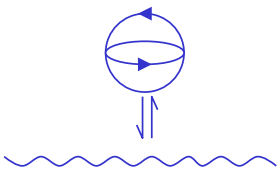


Fig. 1.2 Maturana and Varela's ideograms to denote structurally-determined autopoietic and operationally-closed systems. The diagram (denotes an organizationally-closed autonomous system with a central nervous system. This system is capable of development by means of self-modification of its nervous system, so that it can accommodate a much larger space of effective system action.

1.3.2 Enaction and Knowledge

Let us now move on to discuss in a little more detail the nature of the knowledge that an enactive cognitive system constructs. This knowledge is built on sensorimotor associations, achieved initially by exploration of what the world offers. However, this is only the beginning. The enactive system uses the knowledge gained to form new knowledge which is then subjected to empirical validation to see whether or not it is warranted (we, as enactive beings, imagine many things but not everything we imagine is plausible or corresponds well with reality). One of the key issues in cognition is the importance of internal simulation in accelerating the scaffolding of this early developmentally-acquired sensorimotor knowledge to provide a means to predict future events, to reconstruct or explain observed events (constructing a causal chain leading to that event), or to imagine new events [132, 144, 337]. Naturally, there is a need to focus on (re-)grounding predicted, explained, or imagined events in experience so that the system can *do* something new and interact with the environment in a new way. If the cognitive system wishes or needs to share this knowledge with other cognitive systems or communicate with other cognitive systems, it will only be possible if they have shared a common history of experiences and if they have a similar phylogeny and a compatible ontogeny.

In other words, the meaning of the knowledge that is shared is negotiated and agreed by consensus through interaction.

When there are two or more cognitive agents involved, interaction is a shared activity in which the actions of each agent influence the actions of the other agents engaged in the same interaction, resulting in a mutually constructed pattern of shared behaviour [275]. Again, Maturana and Varela introduce a succinct diagrammatic way of of conveying this coupling between cognitive agent and the development it engenders [238]: see Fig. 1.3. Such mutually-constructed patterns of complementary behaviour is also emphasized in Clark's notion of joint action [64]. Thus, explicit meaning is not necessary for anything to be communicated in an interaction, it is simply important that the agents are mutually engaged in a sequence of actions. Meaning emerges through shared consensual experience mediated by interation. The research programme encapsulated in this roadmap is based on this foundational principle of interaction. The developmental progress of imitation follows tightly that of the development of other interactive and communicative skills,

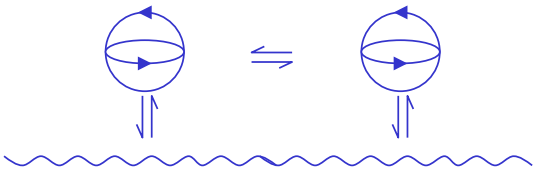


Fig. 1.3 Maturana and Varela's ideogram to denote the development engendered by interaction between cognitive systems

such as joint attention, turn taking and language [268, 350, 377]. Imitation is one of the key stages in the development of more advanced cognitive capabilities.

1.3.3 Phylogeny and Ontogeny: The Complementarity of Structural Determination and Development

Let us summarize: enaction entails two complementary processes: (a) phylogenetically-dependent structural determination, i.e. the preservation of autonomy by a process of self-organization which determines the relevance and meaning of the system's interactions, and (b) ontogenesis, i.e. the increase in the system's predictive capacity and the enlargement of its action repertoire through a process of generative model construction by which the system develops its understanding of the world in which is it embedded. Ontogenesis results in development: the generation of new couplings effected by the self-modification of the system's own state, specifically its central nervous system. This complementarity of structural determination — phylogeny — and development — ontogeny — is crucial. Cognition is the result of a developmental process through which the system becomes progressively more skilled and acquires the ability to understand events, contexts, and actions, initially dealing with immediate situations and increasingly acquiring a predictive or prospective capability. Prediction, or anticipation, is one of the two hallmarks of cognition, the second being the ability to learn new knowledge by making sense of its interactions with the world around it and, in the process, enlarging its repertoire of effective actions. Both anticipation and sense-making are the direct result of the developmental process. This dependency on exploration and development is one of the reasons why the artificial cognitive system requires a rich sensory-motor interface with the environment.

1.4 Embodiment: The Requirements and Consequences of Action

Cognitive systems as described above are intrinsically embodied and embedded in their environment in a situated historical developmental context [370]. Furthermore, as we have already noted, the system's physical embodiment plays a direct constitutive role in the cognitive process [383, 204, 111]. But what exactly is it to be embodied? One form of embodiment, and clearly the type envisaged by proponents of the enactive systems approach to cognition, is a physically-active body capable of moving in space, manipulating its environment, altering the state of the environment, and experiencing the physical forces associated with that manipulation [367]. But there are other forms of embodiment. Ziemke introduced a framework to characterise five different types of embodiment [413, 414]:

1. *Structural coupling* between agent and environment in the sense that a system can be perturbed by its environment and can in turn perturb its environment.
2. *Historical embodiment* as a result of a history of structural coupling;

3. *Physical embodiment* in a structure that is capable of forcible action;
4. *Organismoid embodiment*, i.e. organism-like bodily form (e.g. humanoid or rat-like robots);
5. *Organismic embodiment* of autopoietic living systems.

These five types are increasingly more restrictive. Structural coupling entails only that the system can influence and be influenced by the physical world. Historical embodiment adds the incorporation of a history of structural coupling to this level of physical interaction so that past interactions shape the embodiment. Physical embodiment is most closely allied to conventional robot systems, with organismoid embodiment adding the constraint that the robot morphology is modelled on specific natural species or some feature of natural species. Organismic embodiment corresponds to living beings.

To repeat again, the fundamental idea underpinning embodiment is that the morphology of the system is actually a key component of the systems dynamics. The morphology of the cognitive system not only matters, it is a constitutive part of the cognitive process and cognitive development depends on and is shaped by the form of the embodiment. There is, however, an important consequence of this. In a system that only satisfies the minimal requirements of embodiment, there is no guarantee that the resultant cognitive behaviour will be in any way consistent with human models or preconceptions of cognitive behaviour. Of course, this may be quite acceptable, as long as the system performs its task adequately. However, if we want to ensure compatibility with human cognition, then we have to admit the stronger version of embodiment and adopt a domain of discourse that is the same as the one in which we live: one that involves physical movement, forcible manipulation, and exploration, and perhaps even human form [50]. Why? Because when two cognitive systems interact or couple, the shared consensus of meaning — the systems' common epistemology — will only be semantically similar (have similar meaning) if the experiences of the two systems are compatible: phylogenetically, ontogenetically, and morphologically consistent [237]. Consequently, the approach to cognition we are advocating here requires that the cognitive system be embodied in a very specific sense: that it should lie in the organismoid space of embodied cognitive systems and, further still, that it should lie in the humanoid subspace of the organismoid space.

Apart from the morphology and phylogeny of the cognitive system, this also has strong implications for the development of the cognitive system. Specifically, the ontogeny of the system must follow the development of natural (human) systems. We will deal with this in considerable depth in Chaps. 3 and 6 but it should be noted here that this development follows a general path that begins with actions that are immediate and have minimal prospection, and progresses to much more complex actions that bring forth much more prospective cognitive capabilities. This involves the development of perception-action coordination, beginning with head-eye-hand coordination, progressing through manual and bi-manual manipulation, and extending to more prospective couplings involving inter-agent interaction, imitation, and (gestural) communication. As we will see in Chap. 3, this development occurs in both the innate skills with which phylogeny equips the system and in the acquisition

of new skills that are acquired as part of the ontgenetic development of the systems. As we have noted already, it is the ontogenetic development which provides for the greater prospective abilities of cognitive systems.

1.5 Challenges

The adoption of an enactive approach to cognitive systems poses many challenges. We highlight just a few in the following.

The first challenge is the identification of the phylogenetic configuration and the ontogenetic processes. Phylogeny — the evolution of the system configuration from generation to generation — determines the sensory-motor capabilities that a system is configured with at the outset and that facilitate the system's innate behaviours. Ontogenetic development — the adaptation and learning of the system during its lifetime — gives rise to the cognitive capabilities that we seek. To enable development, we must somehow identify a minimal phylogenetic state of the system. In practice, this means that we must identify and effect perceptuo-motor capabilities for the minimal behaviours that ontogenetic development will subsequently build on to achieve cognitive behaviour.

The requirements of real-time synchronous system-environment coupling and historical, situated, and embodied development pose another challenge. Specifically, the maximum rate of ontogenetic development is constrained by the speed of coupling (i.e. the interaction) and not by the speed at which internal processing can occur [400]. Natural cognitive systems have a learning cycle measured in weeks, months, and years and, while it might be possible to condense these into minutes and hours for an artificial system because of increases in the rate of internal adaptation and change, it cannot be reduced below the time-scale of the interaction. You cannot short-circuit ontogenetic development because it is the agent's own experience that defines its cognitive understanding of the world in which it is embedded. This has serious implications for the degree of cognitive development we can practically expect of these systems.

Development implies the progressive acquisition of anticipatory capabilities by a system over its lifetime through experiential learning. Development depends crucially on the motives which underpin the goals of actions. The two most important motives that drive actions and development are social and exploratory. There are at least two exploratory motives, one focussing on the discovery of novelty and regularities in the world, and one focussing on the potential of one's own actions. A challenge that faces all developmental embodied robotic cognitive systems is to model these motivations and their interplay, to identify how they influence action, and thereby build on the system's phylogeny through ontogenesis to develop every-richer cognitive capabilities.

Our goal in this book is to address these issues by borrowing heavily from the neurosciences and from developmental psychology, using them to guide us in identifying the necessary phyologeny, the progression of ontogenetic development, the balance between the two, and the factors that drive the ontogenetic development. We consider these in detail in the next three chapters.

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

perception [32]. Perception is needed both for planning actions and for guiding them toward their goals. However, not only does action rely on perception, it is also a necessary part of the perceptual process. For instance, active touch is required to haptically perceive the form of an object [117]. The hand must move over the object and feel its form, its bumps and its indentations. The clearest example of the necessity of action for functional perception is vision itself. Our visual field consists of a very small fovea surrounded by a large peripheral visual field over which acuity rapidly deteriorates with increasing angular eccentricity. In spite of this, we have the illusion that we see equally clearly over our whole field of vision. A simple experiment shows that this is wrong. If one firmly fixates a word in a text it is hardly possible to even read the neighboring words. The illusion of an equally clear visual field is created by the fact that we move the fovea to every single detail that we want to inspect, by doing this we can inspect it with optimal resolution. The same principles hold for all modes of perceiving. Perception is always characterized by exploratory activities such as looking, listening, sniffing, tasting, and feeling [117]. It is equally true that all actions also have perceptual functions. Locomotion reveals the layout of the environment, manipulation reveals object properties, and social interaction is essential for person perception. One's movements also reveal information about the biomechanics of the body, the forces acting on it and how these change over the execution of a movement. Thus, by necessity, any action also involves perceptual actions.

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

2.2 Prenatal Development

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

2.2.1 Morphological Pre-structuring

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

What is less obvious but equally true is that each of these body parts is a part of a perception-action system that also includes specially designed perceptual and neural mechanisms. The design of the body of any animal, its sensory and perceptual system, its effector system, and indeed its neural system have been tailored to each other for solving specific action problems. The changes in the morphology of the body also include adjustments of the perceptual system to improve extraction of information for controlling specific actions. For instance, the frontal positions of the eyes in primates give access to better information for controlling manual movements. It should be noted, however, that the same evolutionary change decreases the size of the visual field and decreases the ability to quickly detect predators. Thus, in some species, like the deer and the rabbit, the ability to detect predators is optimized by having the eyes positioned to the sides of the head and thereby enlarging the visual field. In other species, like the cat and the monkey, the ability to perform precise actions is optimized by directing the eyes forward and thereby making it possible to coordinate the two visual fields into one. Precise manipulation is greatly facilitated by the evolution of detailed foveal vision, by the ability to precisely converge and accommodate the eyes on the point of interest and track objects over space, and by the evolution of direct cortico-motor-neuronal pathways that makes it possible to control individual finger movements [209].

In lower vertebrates, it often appears as if action systems have evolved independent of each other. Thus the frog seems to possess independent perceptual mechanisms for extracting spatial information needed for catching flies and for negotiating barriers [327]. In higher vertebrates, movement patterns are more flexible and the perceptual skills more versatile. When a new skill evolves, the animal may re-use some of the mechanisms already evolved for other tasks instead of developing completely new ones. This leads to more general mechanisms and more generalized skills. A similar trend seems to be going on in ontogeny. The earliest appearing skills seem more task specific than those appearing later.

2.2.2 Pre-structuring of the Motor System

Simply providing the hardware is not sufficient for establishing a perception-action system. In addition there need to be some initial constraints on the movements produced in order to reduce the many degrees of freedom of the motor system [32].

To facilitate control, the activation of muscles is therefore organized into functional synergies at the beginning of life. Synergies have both facilitating and constraining effects. For instance, the arm and finger movements of newborn infants are organized into extension and flexion synergies that make the arm and the fingers extend and flex together. These synergies simplify the control problem and enable newborn infants to direct movements of their arms in space. However, it prevents the neonate from grasping an object reached for because that would require them to flex the hand around the object while the arm is extended.

Organized movements of the human child are observable from the 9th week of gestation [389]. Within a month the foetus will begin to make organized breathing movements, open and close the mouth, yawn, suck, and swallow (see Fig. 2.1). They will move their arms and hands and turn the head in an organized way. There is evidence, that the foetus moves its hands and legs to touch the walls of the amniotic sack, grasp the umbilical cord, and put the thumb in the mouth (see Fig. 2.2). At 22 but not 18 weeks of gestation the hand movements of a foetus are planned in the sense that those directed to the eye are more smooth, decelerated, than to the ones towards the mouth [416].

A newborn child will perform walking movements under certain conditions. This neonatal walking is organized in a similar way as in other mammals with the toes being lowered ahead of the heels [100]. Neonatal stepping is optimized when an optic flow is presented visually [21]. When awake 3-day-old human infants are vertically positioned above an optic flow their stepping is related to the characteristics

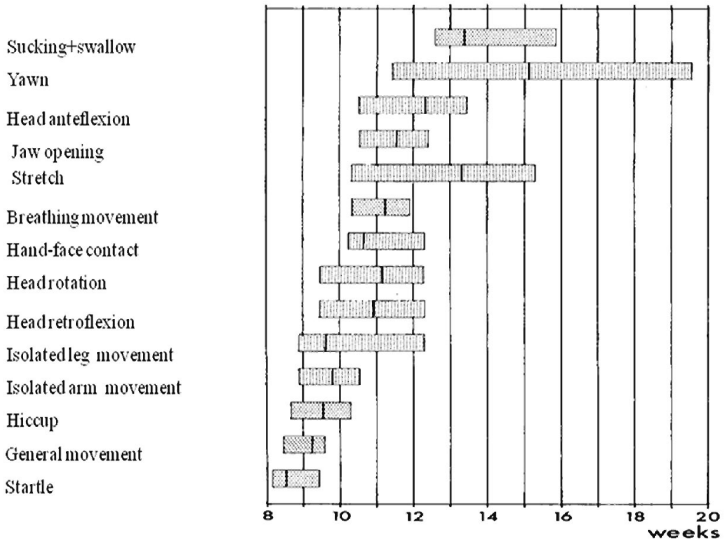


Fig. 2.1 Age intervals when various prenatal behaviour appear. These intervals were determined with ultrasound technology (from [389]).

Fig. 2.2 A 20-week-old fetus who puts the thumb into the mouth (from [274])

of the flow. It is concluded that the visual information of flow direction and velocity influences the leg movements. Other studies have shown that the stepping is influenced by the external condition in which it is performed. When infants' legs were loaded with small weights to simulate normal gains in leg fat, previously stepping infants stopped stepping [369]. Conversely, when infants' legs were submerged in a tank of water to alleviate the effects of gravity, non-stepping infants stepped once again.

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

2.2.3 Pre-structuring of the Perceptual System

Perception also requires some structuring to begin with in order to provide the necessary guidance for action. Infants must be able to perceive speech sounds in order to be ready to produce them. Research in this area has shown that speech perception develops ahead of speech production [247]. Before vision can guide looking, the visual field must be directionally structured and before it can guide object directed action, it must be able to divide up the perceptual field into object defining entities. Although little is known about when these processes of perceptual structuring start to emerge in development, some of the actions performed by newborn infants indicate that object perception is present at birth.

The early structuring of vision is accomplished prenatally and provides a beautiful example of the parsimony of the embryo-genetic process. It may serve as an

example of the more general principles of neural mapping. It is a two-stage process. Both stages of mapping are necessary [229]. The first stage is primarily determined by the genotype and the second stage by the activity of the foetus. First, an abundance of axons originating at the retinal level migrate to the thalamus (the lateral geniculate nucleus) and the superior colliculus under guidance of genetically determined chemical gradients where they will form topographies crudely corresponding to the retinal topography [305]. The resulting projections are, however, too fuzzy for extracting specific information about the world.

At the second stage of the mapping, structured activity at the retinal level will cause connections to be modulated through competitive interactions [229]. Strong connections become strengthened [139] and will successfully compete with the weaker connections for the limited synapse space available. This will transform the initial crude mapping into a detailed one. Spontaneous neural activity at the retinal level ensures that enough structured activity at the retinal is provided to map up the visual system [340]. It is possible that the spontaneous activities of the foetus facilitate the mapping of the visual system. Moving the arms in front of the eyes in the womb produces moving shadows over the eyes that might assist in the mapping of the visual system. In addition, the change in the light level when the arms move in front of the eyes provides information about the contingencies between arm movements and visual input.

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

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

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

2.2.4 Forming Functional Systems

The various constraints set up by phylogeny will selectively sponsor the growth and structuring of pathways in the nervous system that are parts of functional systems which the child needs at birth [8]. As a consequence of this selective, accelerated growth, neonates are prepared to sustain life in their new environment and to explore and adapt to it. Anokhin [8] gives a number of examples of such accelerated growth. For instance, although the facial nerve is an isolated structure, it shows a marked disproportionate maturation of several fibres at birth. The fibres projecting to M. orbicularis oris, providing the most important movement in sucking, are already myelinated and the contacts with the muscle fibers established at a stage when no other facial muscles have such marked organization. Similar accelerated growth can be observed in the medulla oblongata. The parts related to the functional system of sucking are ready to be used, while, for instance the parts that are the source of the frontal branches of the N. Facialis, are just beginning to differentiate. The fact that the morphogenesis of the nervous system primarily follows functional rules rather than structural ones was called “the principle of systemogenesis” by Anokhin [8].

2.3 Core Abilities

2.3.1 Core Knowledge

To facilitate the acquisition of particular kinds of ecologically important knowledge, basic aspects of them are prestructured in human infants. This is valid for the perception of objects and their motions, the perception of geometric relationships and numerosities, and the understanding persons and their actions. Work with other species indicates that these systems have a long evolutionary history. Nevertheless, core knowledge systems are limited in a number of ways: they are domain specific (each system represents only a small subset of the things and events that infants perceive), task specific (each system functions to solve a limited set problems), and encapsulated (each system operates with a fair degree of independence from other cognitive systems) [352]. Knowledge about objects, space, numbers, and people are a few of them.

2.3.1.1 Objects

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

Object perception does accord with principles governing the motions of material bodies: infants divide perceptual arrays into units that move together, that move separately from one another, that tend to maintain their size and shape over motion, and that tend to act upon each other only on contact. To be perceived as an object, there must be well-defined and persistent outer boundaries. A heap of sand, for instance,

is not perceived as an object. These findings suggest that a general representation of object unity and boundaries is interposed between representations of surfaces and representations of objects of familiar kinds [351].

Perceived objects move on continuous and unobstructed paths. When motion carries an object fully out of view, the object is expected to continue on the same path. Baillargeon and associates [19, 2] habituated infants to a tall and a short rabbit moving behind a solid screen. This screen was then replaced by one with a gap at the upper part. The tall rabbit should have appeared in the gap but did not. Five-and-a-half-, three-and-a-half-, and two-and-a-half-month-old infants looked longer at the tall rabbit event suggesting that infants had detected a discrepancy between the expected and the actual motion of the rabbit in that display. When infants visually track an object that disappears temporarily behind another one during its motion they stop tracking at the border of disappearance and shift gaze to a position at the extension of the previous trajectory before the object reappears there [158, 325, 200]. This behavior emerges around 3 months of age [323] and at 4 months it is functionally mature. Then, infants will adjust the latency of moving gaze to the reappearance edge to the velocity of the moving object and the width of the occluder. These behaviors are not rigid, however. If the object does not reappear at the expected location, infants quickly learn a new reappearance location [158, 200]. Kochukhova and Gredebäck [200] found that 6-month-old infants who visually tracked a moving object that disappeared behind an occluder after having moved on a straight path began to expect the object to reappear on a path perpendicular to the original one after this had occurred on only 2 trials. When the object disappears, infants do not shift gaze to the expected reappearance position right away. They rather wait until the object is about to reappear before making a saccade over the occluder [130, 161]. Such behavior is seen consistently from 3-4 months of age [325, 161].

2.3.1.2 Numbers

Young infants have two core knowledge systems related to numbers: one that deals with small, exact numbers of objects and one that deals with approximate numerosities of sets [352, 92]. The knowledge about exact numbers seems to have a limit of 3. Infants' discriminate 1 vs. 2 and 2 vs. 3 reliably but not any higher numbers. The exact number concept is not dependent on modality. Infants prefer to look at an array of objects that corresponds in number to a sequence of sounds. Three tones and 3 objects are perceived as equal in this respect. [358]. Infants also have the ability to add these small numbers. Wynn [408] found that when one doll was hidden behind an occluder and another doll was hidden there as well, infants expected two dolls to be present when the occluder was removed. Thus, the exact core number concept seems to have a limit of 3. When 10- and 12-month-old infants were shown crackers being hidden in two different buckets, they choose the one with more crackers up to 3. With any higher numbers the choice was random [93]. When 14-month-old infants saw objects being hidden sequentially in a box and then were able to search for them, they retrieved all of them if the number of objects were 1, 2, or 3. However, when 4 objects were hidden, infants retrieved one of them and then stopped

searching [92]. The approximate number system enables infants to discriminate larger sets of entities. Xu and Spelke [409] found that 6-month-old infants' discriminated numerosities 8 vs. 16 using a habituation paradigm. Infants' numerical discriminations are imprecise and subject to a ratio limit: 6-month-old infants successfully discriminate 8 vs. 16 but fail with 8 vs. 12. Second, numerical discrimination increases in precision over development and adults can discriminate ratios as close as 7:8 [27].

2.3.1.3 Space

Research on animals, including humans, suggest that navigation is based on representations that are dynamic rather than enduring, egocentric rather than geocentric, and limited to a restricted subset of environmental information. Uniquely human forms of navigation build on these representations [391]. The evidence comes from studies of path integration, place recognition, and reorienting based on congruence finding on representations of the shape of the surface layout. Path integration has been found to be one of the primary forms of navigation in insects (see e.g. [266], birds (see e.g. [303], and mammals [110]. Like other animals, humans can return to the origin of a path and travel to familiar locations along novel paths [212]. When asked to point to objects in familiar locations while moving around blindfolded, the errors made by subjects accumulate just like they do with path integration in animals. If, however, the subjects were shown just one single beam of light the errors stay small and constant. This shows that the errors were not caused by forgetting but rather by disorientation. Like animals, humans orient by recognizing places rather than by forming global representations of scenes. Gillner and Mallot [119] studied how people learn to navigate through a virtual neighbourhood of interconnecting streets furnished with multiple landmarks. Patterns of travel provided evidence that people learn to turn in specific directions at particular places, and that their turning decisions depend on local, view-dependent representations of landmarks. That children use such a geometry-based reorientation system is also suggested by Hermer and Spelke [141]. They studied 1.5- to 2-year-old children who saw a toy hidden in one corner of a rectangular chamber, were then disoriented by turning, and finally released and encouraged to find the toy. In different experiments, the location of the toy was specified by the distinctive color of a single wall or by the presence of a distinctive landmark object. Like rats, children searched reliably and equally at the correct corner and at the geometrically equivalent opposite corner. Their successful use of room geometry showed that they were motivated to perform the task, remembered the object's location, and, like rats, reoriented in accordance with the shape of the surface layout but not by non-geometric landmarks.

2.3.1.4 People

An important part of core knowledge has to do with people. Infants are attracted by other people, endowed with abilities to recognize them and their expressions,

communicate with them, and perceive the goal-directedness of their actions. The motions produced by a moving person are preferred over other motions in young infants. Fox and Daniel [101] demonstrated that 8-week-old infants preferred a point-light walker over dynamic noise or the same configuration inverted in the image. It has even been shown that newborn infants are sensitive to biological motion [345] suggesting that this ability is innate. Intentions and emotions are displayed by elaborate and specific movements, gestures, and sounds that become important to perceive and control. Some of these abilities are already present in newborn infants and reflect their preparedness for social interaction. Neonates are very attracted by people, especially to the sounds, movements, and features of the human face [239, 89]. They have a greater tendency to visually track a schematic face than one where the facial parts are scrambled inside the outer contour [184]. They look longer at a face that directs the eyes straight at them than at one that looks to the side [89]. They also engage in some social interaction and turn-taking that among other things is expressed in their imitation of facial gestures [246]. Finally, they perceive and communicate emotions such as pain, hunger and disgust through their innate display systems [403]. These innate dispositions give social interaction a flying start and open up a window for the learning of the more intricate regularities of human social behavior. Parents show a remarkable talent for responding to the infant's signals and turning them into sophisticated forms of social interaction. Rochat and Striano [360] suggested that this "propensity to express empathy through the echoing of affects and feelings in highly scaffolding ways is part of normal parenting and ... the primary source of intersubjectivity".

2.3.1.5 Core Knowledge Systems and the Development of Cognitive Capabilities

Spelke [352, 353] suggests that the core knowledge systems found in infants contribute to later cognitive functioning in two ways. First, core systems continue to exist in older children and adults, giving rise to domain-specific, task-specific, and encapsulated representations like those found in infants. Second, core knowledge systems serve as building blocks for the development of new cognitive skills. When children or adults develop new abilities to use tools, to perform symbolic arithmetic calculations, to read, to navigate by maps and landmarks, or to reason about other people's mental states, they do so in large part by assembling in new ways the representations delivered by their core knowledge systems. Language presumably plays an important role in this process.

2.3.2 Core Motives

The development of an autonomic organism is crucially dependent on motives. They define the goals of actions and provide the energy for getting there. The two most important motives that drive actions and thus development are social and explorative.

They both function from birth and provide the driving force for action throughout life.

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

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

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

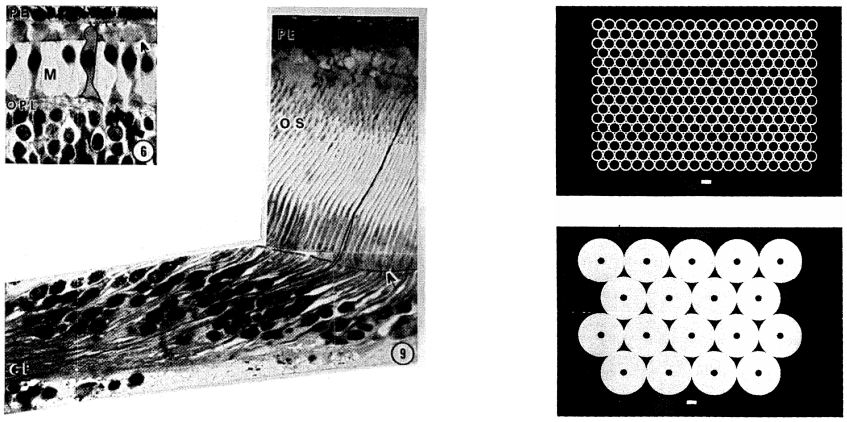
Two processes of perceptual development can be distinguished. The first one is a spontaneous perceptual learning process that has to do with the detection of structure

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]).

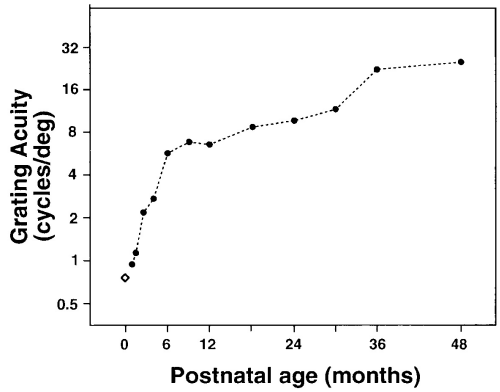


Fig. 3.2 Changes in grating acuity between birth and 48 months when measured by preferential looking (from [240]).

that the infant has an innate sensitivity for certain retinal patterns or templates or that certain retinal patterns are learnt shortly after birth because the pattern of excitations will not be the same over development. As a result of the changes occurring on the retina and in the ganglions further back, the visual acuity improves dramatically during the first few months of life. As can be seen from Fig. 3.2, the acuity at 5 months of age is about a quarter of the adult acuity.

Several of the basic cortical visual functions are not available at birth but mature during the first half year of life. Thus, colour perception is deficient at birth but functions from about 1 month of age. Certain aspects of motion perception are available at birth and are then processed subcortically. When the projection of a stimulus moves over the retina, the light level will change momentarily for individual retinal positions. Certain receptors are sensitive to this light level flickering [9]. They thus react to visual motion but without a directional component. Sensitivity to motion direction requires that flickering receptors are sequentially activated. Newborn infants are sensitive to flickering and attracted by it. Therefore they move the eyes to a position in the visual field that exhibits motion. However, they cannot yet perceive directed motion and track a moving stimulus.

Von Hofsten and Rosander [164, 165] recorded eye and head movements in unrestrained 1- to 5-month-old infants as they tracked a happy face moving sinusoidally back and forth in front of them. They found that the improvement in smooth pursuit tracking was very rapid and consistent between individual subjects. Smooth pursuit starts to improve around 6 weeks of age and attain adult levels from around 14 weeks. The ability to discriminate motion direction emerges during the same period (see [14]). ERP studies show that the MT-MST area is engaged in motion processing at least from around 8 weeks of age and is fully functional from about 14-18 weeks [323].

3.2.1 Space Perception

Visual space perception relies primarily on binocular information, motion information, and a whole set of monocular cues that induce depth in pictures. They all develop during the first year of life but at different schedules [192]. Let's first consider motion as information for depth. There is such information in the expansion of the retinal projection of an approaching object, the motion parallax on the retina when the subject moves, and the accretion-deletion of object structure at the edge of an occluding object when one object moves behind another. The earliest signs of sensitivity to space from motion come from studies of looming. Reliable effects of increased blinking to approaching displays have been found in several studies with infants from less than a month on [411, 269]. Kayed & van der Meer [191] found that the youngest infants blinked when the virtual object reached a threshold visual angle, while older ones geared their blinks to the virtual object's time-to-collision. The shift did not occur until at around 6 months of age. This indicates that although young infants perceive that an object is approaching, they cannot evaluate so well when it is going to hit them.

Sensitivity to motion parallax was demonstrated in 3-month-old infants by von Hofsten, Kellman, and Putaansuu [160]. They showed infants an array of 3 vertical rods in a horizontal row, perpendicular to the line of sight. When the infant moved laterally in front of these rods, the middle one moved in phase with the infant. Afterwards they were tested with 3 stationary rods with the middle one either aligned with the other ones (as in the original display) or displaced backwards to an extent corresponding to the contingent motion. When the velocity of the contingent rod was $0.32^\circ/s$ (visual angle), the infants looked significantly more at the 3 aligned stationary rods than at the display where the middle rod was displaced backwards to an extent corresponding to the contingent velocity. When the contingent velocity was decreased to $0.16^\circ/s$, the looking at the test display did not show any preference. The results are consistent with the idea that young infants utilize small contingent optical changes as information about depth. The results do not uniquely imply this interpretation, however. It might simply be that infants are very sensitive to optical changes contingent on their own motion. These optical changes do appear special in that infants' sensitivity to them exceeded what has been found in other studies of motion sensitivity by almost an order of magnitude (see e.g. [13, 74]).

Binocular depth perception relies on two mechanisms – sensitivity to the convergence of the eyes and sensitivity to binocular disparity. Convergence gives absolute distance and disparity relative depth to objects in the surrounding. By 1 month of age, convergence operates accurately for distances beyond 20 cm ([135]. Von Hofsten [145] showed that by 5 months of age, infants use convergence information about distance when programming reaching movements (see Fig. 3.3). It is possible, however, that convergence is used much earlier in life, maybe even at birth. Kellman, von Hofsten, van der Walle & Condry [193] showed displays to young infants that contained several stationary objects and one that moved contingent on the movement of the infant. In order to perceive which object was moving, the infant had to correctly perceive the distance to it. 8-week-old infants consistently

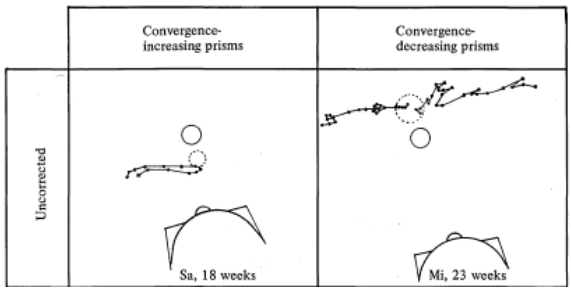


Fig. 3.3 Object reaches performed with convergence increasing and convergence decreasing prism spectacles, transcribed from video-records. The time interval between each plotted dot is 100 ms. The real object is indicated by a circle with a solid outline, and the virtual object is indicated by a circle with a broken outline (from [148]).

discriminated displays containing a moving object from those with only stationary ones. As 8-week-olds have been found to process binocular disparity information [91], this is most probably responsible for the effect. Other signs of binocular depth perception have been observed in 8-week-old infants, but many infants first show sensitivity a month or so later. Birch, Gwiazda, & Held [42] found reliable preferences at 12 weeks of age for crossed disparities and at 17 weeks of age for uncrossed. Improvement in stereoscopic acuity once it appears is quite rapid. Binocular sensitivity improves from 60' visual angle to less than 1' in just a few weeks [140]. In this respect, the ability to process binocular depth show a parallel development relative to the ability to process visual motion. Indicators of perceived motion show that this ability also emerges within just a few weeks [165, 14].

The development of sensitivity to pictorial depth information comes primarily from studies by Yonas and colleagues (see [410]). Many of these studies used reaching as dependent measure. They systematically examined the different depth cues, including linear perspective, familiar size, interposition, and shading. The results are quite consistent and suggest that infants do not utilize pictorial depth cues to guide reaching until they are 6-7 months old. It is possible that several of the pictorial depth cues originate from dynamic situations. For instance interposition refers to the cue that an object that is partly hidden by another is perceived to continue behind it. Granrud & Yonas [127] found that 7-month-old but not 5-month-old infants utilize this cue when reaching for objects. The dynamic version of this cue is the gradual accretion and deletion of object texture as one object goes behind another. 5-month-old infants reliably use this information in predicting when and where an object that disappears behind another will reappear on the other side [33].

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

3.2.2 Object Perception

The rules by which infants perceive objects as separate entities are similar to the ones used by adults. Objects are defined by outer boundaries and inner unity that are preserved over time. To be perceived as an object, there must be well-defined and persistent outer boundaries. A heap of sand, for instance, is not perceived as an object. This suggests that a general representation of object unity and boundaries is interposed between representations of surfaces and representations of objects of familiar kinds [351].

To define the outer boundaries and the inner unity, motion information is relatively more important than static information early in life. Infants divide perceptual arrays into units that move together, that move separately from one another, that tend to maintain their size and shape over motion, and that tend to act upon each other only on contact. Two units that move relative to each other are perceived as separate objects and two units that move together are perceived as a single object [166, 355]. Units that are separated in depth, thus creating relative retinal motion as the subject move, are also perceived as separate objects. If only parts of an object are visible and a nearer object occludes the space between them, the parts are still perceived as belonging to one object if the occluded object moves or the subject moves. Kellman and Spelke [194] found that object pieces protruding on each side of an occluder were not perceived as belonging to the same object if they were stationary. However, if the pieces moved with a common motion along the occluder, 3-month-old infants perceived them to belong together and to be connected behind the occluder. This was the case both when the pieces showed good continuation behind the occluder, such as being parts of a single rod, and when they were totally dissimilar. Smith et al. [349] found that when the pieces protruding from behind the occluder were misaligned relative to each other, common motion had a somewhat weaker binding effect. They concluded that alignment information could enhance perception of object unity either by serving directly as information for unity or by optimizing the detectability of motion-carried information for unity. Van de Walle & Spelke showed 5-month-old infants objects whose centre was fully occluded and whose ends were visible only in succession. Infants perceived this object as one connected whole when the ends of the object underwent a common motion but not when the ends were stationary.

Static object information such as good form, surface colour and texture similarity are much less import as determinants of object unity and boundaries in young infants. Spelke et al. [356] presented adults and infants with simple but unfamiliar displays in which texture similarity, good form, and good continuation either specified one object or two objects. Object perception was assessed by a verbal rating method in the adults and by a preferential looking method in the infants. The Gestalt relations appeared to influence the adults' perceptions strongly. However, the relations appeared to have no effect on the perceptions of 3-month-old infants and only weak effects on the perceptions of 5-month-old and 9-month-old infants. This suggests that motion information dominates infants' perception of objects. Three-month-old infants perceive surfaces in accord with the cohesion principle [356]. Presented with

an array of adjacent surfaces, they perceive a connected body that maintains its connectedness as it moves. These principles apply equally to familiar and unfamiliar forms. Developmental changes in object perception occur only slowly towards a more mature mode where the gestalt principles of good form, surface colour and texture similarity play a more important role.

Colour contributes to the identification of objects at the end of the first year of life. Wilcox et al. [398] found that multi-modal exploration of objects (visual and tactile), but not unimodal (visual only) exploration of objects, prior to an individuation task increased 11-month-old infants sensitivity to colour differences.

3.3 Acquiring Predictive Control

If mastery of actions relies on the perception and knowledge of upcoming events, then the development of actions has to do with acquiring systems for handling such information. It has to do with anticipating both ones own posture and movements, and future events in the world. For every mode of action that develops, new prospective problems of movement construction arise and it takes time to acquire ways to solve them. The knowledge gathered through systematic exploration of a task is structured into a frame of reference for action that makes planning possible. This is the basis of skill. The importance of practice and repetition is not to stamp in patterns of movement or achieve an immutable program, but rather to encourage the functional organization of action systems [302]. These principles will be exemplified with four different modes of action: posture and locomotion, looking, reaching and manipulation, and social skills.

3.3.1 Development of Posture and Locomotion

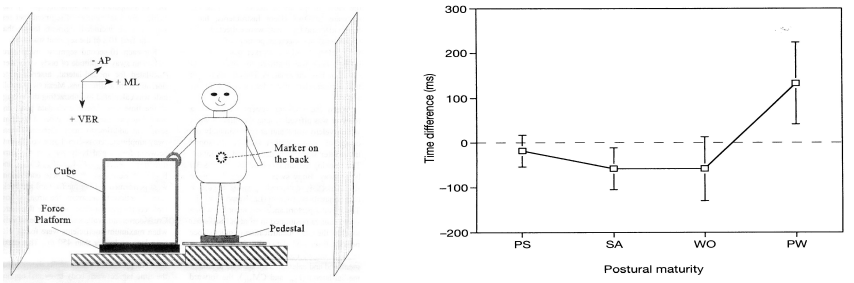
Basic orientation is a prerequisite for any other functional activity [117, 302] and purposeful movements are not possible without it. This includes balancing the body relative to gravity and maintaining a stable orientation relative to the environment. As Reed [302] states, “maintenance of posture in the real world involves much more than simply holding part of the body steady; it is maintaining a set of invariant activities while allowing other activities to vary” (p. 88). Gravity gives a basic frame of reference for such orientational stability and almost all animals have a specialized mechanism for sensing gravity (in humans it is the otoliths). In addition, vision provides excellent orientational information as does proprioception. The contribution of vision is crucial for supporting balance prospectively.

Gravity is also a potent force and when body equilibrium is disturbed, posture becomes quickly uncontrollable. Therefore, any reaction to a balance threat has to be very fast and automatic. Several reflexes have been identified that serve that purpose. For instance, when one slips, a series of fast automatic responses are elicited that serve the purpose of regaining balance. Postural reflexes, however, are insufficient to maintain continuous control of balance during action. They typically interrupt action. Disturbances to balance are better handled in a prospective way, because

if the disturbance can be foreseen there is no need for an emergency reaction and ongoing actions can continue. Another threat to balance is one's own movements. When a body part is moved, the inertia created by the movement will push the body out of equilibrium if nothing is done about it. The movement will also shift the point of equilibrium and that will also disturb balance. Therefore, the effects of one's own movements must be foreseen and prepared in order to maintain ongoing activity.

At around 3 months, infants show the first signs of being able to actively control gravity. When in a prone position they will lift their head and look around. To hold the head steadily, its sway must be correctly perceived and used to control head posture. Such control seems to be attained over the first few weeks of head lifting. The next step in mastering postural control is controlling the sitting posture. This is normally accomplished around age 6-7 months and requires the child to control the sway of both head and trunk in relation to each other. This could be accomplished in a large number of ways because many different muscle groups affect the sitting posture. Woollacott, Debu, and Mowatt [405] found that infants did not show a consistent postural response synergy while sitting until around 8 months of age. Hadders-Algra, Brogren, and Forssberg, [133] tested 5- to 10-month-old infants' postural adjustments when sitting on a platform and being subjected to slow and fast-forward and backward displacements. They found that from the youngest age onwards rather variable, but direction specific muscle activation patterns were present. With increasing age the variation in muscle activation pattern decreased resulting in a selection of the most competent patterns. Barela et al. [22] examined whether there is any developmental change in the coupling between visual information and trunk sway in infants as they acquire the sitting position. Six-, 7-, 8-, and 9-month-olds sat inside a moving room that oscillated back and forward at frequencies of 0.2 and 0.5 Hz. Relative phase showed that at 0.2 Hz, infants were swaying with no lag but at 0.5 Hz they were lagging the room. The results showed that the coupling between visual information and trunk sway in infants varies with the visual stimulus but does not change as infants acquire the sitting position.

In upright stance, the body acts as a standing pendulum. The natural sway frequency of a pendulum is inversely proportional to the square root of its length. This means that the balancing task is much more difficult for a child than for an adult. For instance, a child who is only half the size of an adult will sway with a frequency which is 40 percent higher than that of the adult and will consequently have 40 percent less time in which to react to balance disturbances. In other words, when, by the end of the first year, infants start to be able to stand independently they have mastered a balance problem more difficult than at any time later in life. Barela, Jeka & Clark [23] made a very nice demonstration of the development of predictive control of standing posture. The infants simply stood and held a handrail. The forces that the subject applied to the rail and his/her sway were simultaneously recorded. Four groups of infants were studied according to their postural maturity: prestanding, standing alone, walk onset (> 3 steps), and post walking (walking for more than 1.5 months). The results show that for the first 3 groups the forces applied to the rail lagged the sway but for the post walking infants, the forces preceded the sway (see Fig. 3.4).



(a) (b)

Fig. 3.4 (a) The experimental setup for studying the interaction between sway and forces applied to a reference cube. (b) The time divergence between sway and corrective forces applied to the cube. PS: pre-standing, SA: Stand alone, WO: walking onset (> 3 steps) PW: post walking (> 1.5 months of walking). Only when children had been walking for more than 1.5 months did they apply forces to adjust balance consistently ahead of the sway (from [23]).

Vision is quite superior in detecting small body displacements, and with it, the subject can be more efficient in using prospective control for controlling body sway. Lee and Aronsson (Lee and Aronsson, 1974) showed that infants who have just attained upright stance are quite sensitive to peripheral visual information for body displacement. They positioned standing infants in a room with movable walls and ceiling (the moving room), and when they moved these surrounding structures, the infants lost their balance in the direction predicted by the visual flow. With more experience of standing, children were not as easily overthrown by the visual flow alone. Bertenthal, Rose, and Bai [34] showed that the sensitivity to visual flow improves over the months after upright stance has been achieved. Visual information is especially important for dynamic postural control, that is, when maintaining balance while moving around. Fraiberg [102] found that in a sample of blind children, 90 percent were delayed past the upper limits of sighted children as given by Bayley [29] when walking independently across a room.

Special demands are associated with balance control during bodily activities. In order to maintain balance during limb movements, the subject must know about the contingencies between the limb movements, the reactive forces that arise during movement, and the displacement of the point of gravity. Adults seem to counteract disturbances to the postural system in a precise way ahead of time. Von Hofsten and Woollacott [168] found such anticipatory adjustments of the trunk in 9-month-old infants reaching for an object in front of them while balancing the trunk. Witherington et al. [402] examined the timing of activation of the gastrocnemius muscles when standing infants pulled a drawer that resisted pulling by a weight attached to it. Activation of this muscle counteracts the tendency to fall forward during pulling but only if is activated slightly ahead of the pull. Adults activate the gastrocnemious muscles 50 ms before the arm starts pulling. Witherington et al. [402] found that infants activated this muscle ahead of pulling to an increasing extent between 10

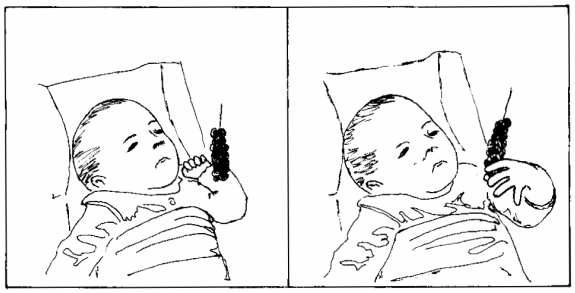


Fig. 3.5 A 1-week-old infant who looks at an object and reaches for it. Note that at the end of the approach, the hand is open. Drawings transcribed from video records (from [155]).

and 17 months of age. The timing, however, was not as precise as in adults. The activation started more like a quarter to half a second before the actual pull. The emergence of independent walking coincided with marked increases in anticipatory postural adjustments of the gastrocnemius muscle relative to pull onset.

Because of its central role in movement production, postural control becomes a limiting factor in motor development. If the infant is given active postural support, goal directed reaching could be observed at an earlier age than is otherwise possible. For instance, the neonatal reaching observed by von Hofsten [148] was performed by properly supported infants (see Fig. 3.5). For these reasons, development of reaching and other motor skills should be studied in the context of posture. However, there are only few studies that have seriously considered the influence of such contextual factors. Rochat and associates [321, 322] showed that the onset of self-sitting made infants transfer from two-handed to one-handed reaching. They suggested that this was because the newly attained posture could be easily disturbed and that two-handed reaching was more threatening to balance than one-handed. Rochat also observed that when infants sitting independently reached forward with one hand, the other one often moved backwards to preserve the point of equilibrium.

Before the onset of bipedal walking two types of locomotions are observed in infants: crawling and cruising (see Fig. 3.6). The standard crawling with knees and hands is the most common type. However, the alternatives (locomotion with slithering on the belly or sitting) are practiced more in homes that have polished floors than in homes with rugged carpets. For the later alternative, the classic or standard crawling is the most common type of locomotion. Recently, it has been shown [308] that standard crawling shares most of the basic principles of other vertebrate quadruped gaits. In a study by Haehl et al. [134], it is suggested that cruising represents an important transition from quadruped to bipedal locomotion. Using support the infant learns to control the trunk and consequently improving the postural control.

During the first year of independent walking, toddlers improve their gait kinematics, master the postural instability, and the pendulum mechanism of walking

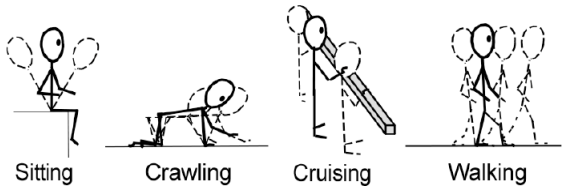


Fig. 3.6 The four postural control systems in their typical order of development. Each posture denotes a distinct problem space that requires unique strategies for obtaining relevant perceptual information, keeping balance, and locomoting through the environment. Adapted with permission from [1].

[181, 248]. One important parameter is head control. Ledebt and Wiener-Vacher [219] conclude that head stabilization in space is achieved during the first weeks of independent walking. During the first year of independent walking, the degree of synchronization between head rotations in the pitch plane and vertical translations increases. Another parameter, reflecting balance control, is step length. New walkers have very short lengths (approx. 12 cm), and with experience these increase ahead of step velocity (25 cm, and 25 to 80 cm/s respectively) [18]. Mastering the ability of bipedal walking is evidently a process of both learning and development. A key question is how a changing environment as well as bodily changes will challenge the infants' control of locomotion. Berger and Adolph [18] state that "the ability to detect affordances lies at the heart of adaptive locomotion". They found, for example, that after 10 weeks of experience, infants geared their locomotor decisions to the possibilities for action. Ivanenko et al. [182] concluded that idiosyncratic features in newly walking toddlers do not simply result from undeveloped balance control but may represent an innate kinematic template of stepping and they summarized different theories for neural control of movements: dynamic systems theory, neuronal group selection, growth and environment.

Recently, two studies have focused on neurophysiological and behaviour evidence for how learning takes place. Sanefuji et al. [331] presented crawlers or walkers with point-light displays of similar actions. They found that crawlers preferred to look at crawling infants, and walkers at walking infants. It was concluded that transformations in the sensory-motor domain are represented similar to those in the physical-visual one, thus supporting a mirror neuron function. This is further demonstrated in a study of van Elk et al. [81]. They measured mu-suppression in EEG for crawlers and walkers when they observed similar videos of crawlers and walkers. The result was that the observation of crawling gave more mu suppression in crawlers, and the observation of walkers induced more mu rhythm suppression in walkers.

3.3.2 Development of Looking

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

3.3.2.1 Shifting Gaze

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

3.3.2.2 Tracking Eye Movements

Several studies on eye movements indicate that newborn infants have only limited ability to track a moving target smoothly. Dayton and Jones [75] found that neonates pursued a wide angle visual display with smooth eye movements but the eye movements became rather jerky for a "small" target. These results were supported by several other studies [43, 202, 12]. Rosander and von Hofsten [324] also found that 1-month-old infants and younger tracked a large moving vertical grating in a smoother way than a small moving target. However, when the saccades were eliminated from the records the residual smooth tracking did not differ for the two targets. In other words, the reason why the tracking of a small target looks jerky is because infants make frequent catch-up saccades in an effort to be on target which they do not need with a large target. The reason is simple. With a wide-field pattern of vertical stripes, the eyes are always on the target, however they move.

From about 6 weeks of age, the smooth part of the tracking improves rapidly. This was first observed both by Dayton and Jones [75] and Aslin [12]. Von Hofsten and Rosander [164, 165] recorded eye and head movements in unrestrained 1- to 5-month-old infants as they tracked a happy face moving sinusoidally back and forth in front of them. They found that the improvement in smooth pursuit tracking was very rapid and consistent between individual subjects. Smooth pursuit starts to improve around 6 weeks of age and attain adult levels from around 14 weeks. The effect of target velocity depended on age. At 2 months of age the proportion of smooth pursuit in the slowest condition (0.2 Hz and 10 deg. amplitude) was almost twice as high as it was in the fastest condition (0.4 Hz and 20 deg. amplitude). At 4 months of age, the proportion of smooth pursuit was high in all conditions and approached adult values.

In order to stabilize gaze on a moving object during tracking, the smooth pursuit must anticipate its motion. Two such predictive processes have been observed in adult visual tracking [282]. One uses the just seen motion to predict what will happen next through a process of extrapolation. Such predictions are in accordance with inertia which presumes that a motion with a certain speed and direction will continue with the same speed and in the same direction unless it is affected by a force in which case the motion will change gradually like in a sinusoidal motion. The extrapolation process is important for predicting object motion over small time windows but it cannot handle prediction over larger time frames. The other predictive process relies on rules: that certain things shall happen at certain times. An abruptly changing motion cannot be extrapolated because the changes do not reveal themselves in the just seen motion. Triangular motion is such a case. The object moves forward with constant velocity and at regular times the motion abruptly reverses. Prediction of these reversals cannot be accomplished through extrapolation but a rule stating the periodicity of the reversals can do it. In order to investigate the development of these two predictive processes, von Hofsten and Rosander [165] studied visual tracking of sinusoidal and triangular motion functions. They found that tracking was well timed in the case of the sinusoidal motion from 2 months of age but the tracking of the triangular motion lagged the target with about 200 ms. At

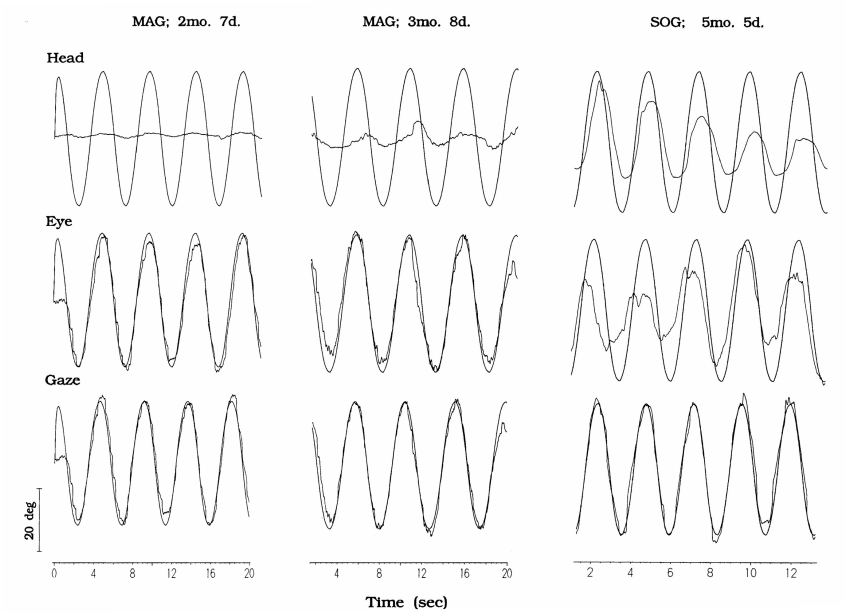


Fig. 3.7 The involvement of eye and head in the tracking of a sinusidally moving object. At all three ages, eye and head supplement each other in the tracking of the object. At the older age, ages the head is more involved (from [165]).

5 months of age, the lagging decreased somewhat indicating that the subjects started to have an idea of the periodicity but most of the lagging remained.

Von Hofsten & Rosander [165] found that the amplitude of head tracking increased very much between 3 and 5 months of age. At 5 months the amplitude of the head tracking was sometimes as large as the amplitude of the object motion (see Fig. 3.7). The problem was that the head still lagged the target at that age (1/3 sec. or more). In order to stabilize gaze on the target, the eyes must then lead. This creates a phase differences between the eye and head tracking that may be so large that the eye tracking and the head tracking counteract each other. Instead of contributing to stabilizing gaze on the fixated moving object, head tracking may then deteriorate gaze stabilization. In fact, the task would be much simpler if the head had not moved at all. The reason why infants persisted in engaging the head can only be because they are internally motivated to do so. Just as in the early development of reaching this is an expression of important developmental foresight because eventually, the ability to engage the head will result in much more flexible tracking skills.

3.3.2.3 Compensatory Gaze Adjustments

Both visual and vestibular mechanisms operate to compensate for head movements unrelated to fixation. The visual one aims at stabilizing gaze on the optic array by minimizing retinal slip while the vestibular one aims at stabilizing gaze in space. The visual mechanism is designed to work at slow optical changes and its performance begins to deteriorate at frequencies above 0.6 Hz [31, 175]. The vestibular mechanism functions most optimally above 1 Hz where the gain approaches unity and the phase lag approaches zero [24]. Head movements unrelated to visual tracking are generally faster and more dynamic than the tracking itself and the eye movements that compensate for those head movements are predominantly guided by vestibular information. This mode of control functions at one month (see Fig. 3.8).

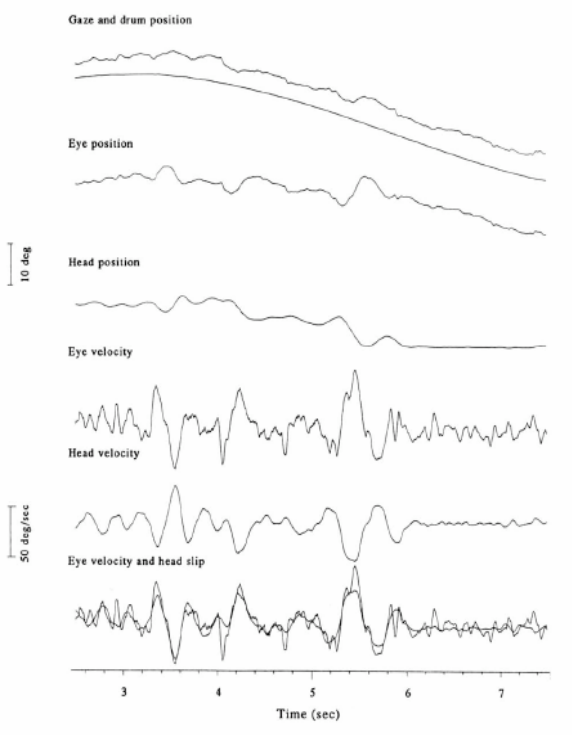


Fig. 3.8 The compensatory character of eye and head movements in the tracking of a moving large target. The two upper curves show that the target and gaze positions are well adjusted to each other. It can also be seen that head position is not so well adjusted to target position but that eye position compensate for the perturbations caused by the head (from [164]).

3.3.3 Development of Reaching and Manipulation

3.3.3.1 Reaching

Visual control of the arm is present at birth [148, 245, 243]. Infants can also move the fingers in a differentiated way, but they cannot control them in grasping or manipulating objects (see Fig. 3.3). Both arm movements and finger movements are governed by global extension and flexion synergies [150]. When the arm extends the fingers extend too and when the fingers flex the arm also flexes. Von Hofsten [150] found that the hand was either open or opened during the extension of the arm in about 70% of the extended arm movements. The opening of the hand did not seem to be a function of the act of reaching towards the object because the same thing happened when the child extended the arm without looking at the object. This pattern was also observed in young rhesus monkeys by Lawrence and Hopkins [218]. They found that newborn monkeys had difficulties in grasping an object they had reached for and if they had finally closed the hand around it, they had difficulties in releasing it after they had pulled it towards them.

von Hofsten [150] found that the synergistic arm-hand pattern changed dramatically at 2 months of age. The coupling was then broken, and instead of opening the hand, the child had a strong tendency to fist it during the extension of the arm. At the same time the movements became more vigorous and appeared more voluntary, as if the child really tried to attain the object [151]. A few weeks later, the subjects were again observed opening the hand during the extension of the arm but then only when the arm movement was visually directed toward the object. The infants then started to close the hand when it was near the object, suggesting that the global extension-flexion pattern had developed into a differentiated pattern where arm and hand were more independently controlled.

Reaching for stationary objects appears between 12 and 18 weeks [65] and catching moving objects appears at approximately 18 weeks [146, 147]. Just as infants' first eye movements are saccadic and lagging rather than smooth and on-target, their first goal-directed reaches and catches are typically jerky and crooked. The transition from pre-reaching to reaching was studied by Thelen et al. [372]. They found that each infant had its own individual way of moving its arms; some moved them more slowly with rather damped movements and some more vigorously. Overall, the early reaching attempts were characterized by much variability which casts doubt on the notion that early movements are stereotyped. During the transition from pre-reaching to successful reaching and grasping the movements became less variable as the infants came to control the intrinsic dynamics of their arms.

Studies of reaching kinematics [146, 152, 36] show that early reaches are rather segmented in contrast to adult reaches which consist of a single bell-shaped velocity curve. von Hofsten [146] defined movement units as segments of the reach, each consisting of an acceleration and a deceleration phase. Corrections are more pronounced during faster reaches [368]. Movement units and direction changes decrease after a few months until an infant's reaches and catches are made up of only two movement units, the first to bring the hand near the target and the second to grasp it. von Hofsten [153] interpreted this development as reflecting increased

prospectivity of the reaching action. What makes an infant's initial reaches so jerky and crooked? One possibility is that movement units reflect visual corrections for a misaligned arm path. However, infants successfully reach for objects in the dark within a week or two of reaching in the light [65], suggesting that they can use proprioceptive information to guide the reach. Indeed, infants fixate the object and not the hand while reaching for a moving object. They also catch moving objects that glow in the dark [320]. By 9 months, they preorient their hands to grasp objects in the dark [241]. Possibly, younger infants have less ability to anticipate the reactive forces that result from the movement itself [35, 368, 154]. Or, infants may have little motivation for efficient reaching [401] — the functional penalty for extra movement units is low — and might even use variable arm paths to explore the capabilities of their new action system [36, 35]. With age, prospective extrapolations of target motion become less dependent on continuous visual information. By 9 months, infants reach for moving objects on an unobstructed path but inhibit reaching when a barrier blocks the path. Six-month-old infants do not plan reaches for moving objects that are temporarily occluded but wait until the object has reappeared [158]. By 11 months, however, infants catch moving objects as they appear from behind an occluder [245].

In the act of reaching for an object there are several problems that need to be dealt with in advance if the encounter with the object is going to be smooth and efficient. The reaching hand needs to adjust to the orientation, form, and size of the object. The securing of the target must be timed in such a way that the hand starts to close around the target in anticipation of and not as a reaction to encountering the object. Such timing has to be planned and can only occur under visual control. Infants do this from the age they begin to successfully reach for objects around 4-5 months of age [163].

Infants are not just able to aim their reaches toward visible object or the remembered position of an object, they are also able to aim their reaches toward future positions of moving objects [147, 149, 162]. Von Hofsten and Lindhagen [162] found that infants reached successfully for moving objects at the very age they began mastering reaching for stationary ones. Eighteen-week-old infants were found to catch an object moving at 30 cm/sec. Von Hofsten [147] found that the reaches were aimed towards the meeting point with the object and not towards the position where the object was seen at the beginning of the reach. Von Hofsten [149] also found that 8-month-old infants successfully caught an object moving at 120 cm/sec. The initial aiming of these reaches was within a few degrees of the meeting point with the target, and the variable timing error was only around 50 msec. Figure 3.9 shows an infant reaching for an object that suddenly stops. In that case the reach is directed toward the point where the object should have been if it had continued on its path. These studies demonstrate that infants predict the future position of a moving object, but they tell us little about the nature or limits of these predictions. Systematic study of the principles guiding predictive reaching requires manipulation of the spatial as well as the temporal properties of object motion. Von Hofsten et al. [167] did this. Infants were presented with an object that moved into reaching space on four trajectories: two linear trajectories that intersected at the centre of a display and two

Fig. 3.9 An 8-month-old infant who attempts to catch an object that suddenly stops. Upper left: the infant prepares to catch the object. Upper right: The object suddenly stops. Lower left: The infant closes the hands around the position where the object should have been if the motion had continued. Lower Right: The infant looks surprised when discovering the true position of the object.

trajectories containing a sudden turn at the point of intersection. Infants' tracking and reaching provided evidence for an extrapolation of the object motion on linear paths, in accord with the principle of inertia. This tendency was remarkably resistant to counterevidence, for it was observed even after repeated presentations of an object that violated the principle of inertia by spontaneously stopping and moving on a nonlinear path.

By 9 months, infants catch moving objects that get temporarily occluded by anticipating their reappearance [244, 143]. In contrast, 6-month-old infants do not plan reaches for moving objects that are temporarily occluded but wait until the object has reappeared [354]. If instead the object is hidden by darkness, the performance is significantly better [187]. How does this reaching behaviour change over development? Hespos et al. [143] tested predictive reaching for occluded objects in 6- and 9-month-old infants. They found that while there was an increase in the overall number of reaches with increasing age, there were significantly fewer predictive reaches during the occlusion compared to visible trials and no age-related changes in this pattern. They also tested adults with a similar reaching task. Like infants, the adults were most accurate when the target was continuously visible and performance

in darkness trials was significantly better than occlusion trials, providing evidence that there is something specific about occlusion that makes it more difficult than merely lack of visibility.

3.3.3.2 Grasping

In the act of reaching for an object there are several problems that need to be dealt with in advance if the encounter with the object is going to be smooth and efficient. The reaching hand needs to adjust to the orientation, form, and size of the object. The securing of the target must be timed in such a way that the hand starts to close around the target in anticipation of and not as a reaction to encountering the object. Such timing has to be planned and can only occur under visual control. Infants do this from the age they begin to successfully reach for objects around 4-5 months of age [279].

When grasping first emerges, infants may use one as well as both hands. The first grasps are power grasps and engage the whole hand. Soon thereafter, however, the radial part of the hand becomes increasingly important for grasping. Although grasping then still involves the whole hand, it tends to be focused on the two most radial fingers and the thumb. Newell et al. [273] studied 4- to 8-month-old infants as they grasped objects that varied in size and shape. The findings revealed that infants as young as 4 months systematically differentiate grip configurations as a function of the object properties in essentially the same way that 8-month-old-infants do. The difference was that younger 4-month-old infants used the haptic system in additional to the visual system for information pick-up regarding object properties, whereas 8-month-old infants predominantly used information from the visual system alone to differentiate grip configurations according to the object properties. Siddiqui [344] presented 5-, 7-, and 9-month-old infants with objects varying from 0.5 to 14.0 cm in diameter. The findings were similar to Newell et al. (op.cit.) in the sense that 5-month-old infants differentiated grip configurations as a function of object size. The number of grasps involving the two or three most radial digits (thumb, index finger, and long finger) increased greatly over this age span. At 9 months of age these kinds of grasps were 10 times more frequent than at 5 months of age. However, at each age level, when only the two or three most radial digits were used, the reaches were typically directed at the two smallest objects. From around 9–10 months of age, infants begin to grasp objects with finger movements that are relatively independent. The independent control of the fingers is made possible by the maturation of the direct cortico-moto-neuronal pathways [209]. When infants develop such finger control, they are able to grasp very small objects with just the index finger and the thumb in precision grasping.

From the age when infants start to reach for objects they have been found to adjust the orientation of the hand to the orientation of an elongated object reached for [225, 157, 159]. The adjustments are crude to begin with but become more precise with age. However, they are never complete. Around 10–15 deg. are always left to be adjusted after contact. When attempting to catch a rotating rod, [159] found that infants prepare the grasping of the object by adjusting the hand to a future

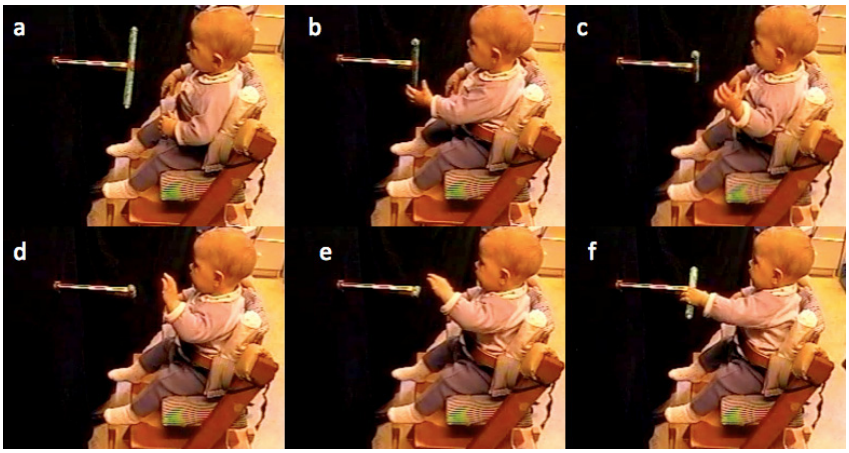


Fig. 3.10 A 9-month-old subject who corrects a reach that was launched too late to accomplish a comfortable end state: (a) the beginning of the reach (b) 1.0 s later, correction starts (c) 0.4 s later, hand is withdrawn. (d) 0.4 s later, a new approach begins. (e) 0.2 s later, the new approach is on its way. (f) 0.4 s later, the grasp is accomplished (from [159]).

orientation of the rod. As they approached the rotating rod from any starting position, they rotated the hand with the rod. If, during the approach, the rotation of the rod exceeded the comfortable rotation of the hand, the infant would flip the hand ahead of time from an over hand to an underhand or vice versa and continue to rotate the hand with the object (see Fig. 3.10).

Adjusting the hand to the size of a target is less crucial. Instead of doing that, it would also be possible to open the hand fully during the approach which would lessen the spatial end point accuracy needed to grasp the object. Adults use this strategy when reaching for an object under time stress [399]. The disadvantage is the additional time it takes to close a fully opened hand relative to a semi-opened hand. Von Hofsten and Rönnqvist [163] found that 9- and 13-month-old infants, but not 5-month-old infants, adjusted the opening of the hand to the size of the object reached for. This is shown in Fig. 3.11. They also monitored the timing of the grasps. For each reach it was determined when the distance between thumb and index finger started to diminish and when the object was encountered. It was found that all the infants including those that just recently had started to reach for objects successfully started to close the hand before the object was encountered. For infants of 9 months and younger the hand started to close rather late during the approach but well before touching the object. For the 13-month-olds, however, the closing of the hand typically started in the middle of the approach. In other words, the hand opened up during the first half of the approach and closed during the second half. Thus, at this age grasping started to become integrated with the reach into one continuous reach-and-grasp act.

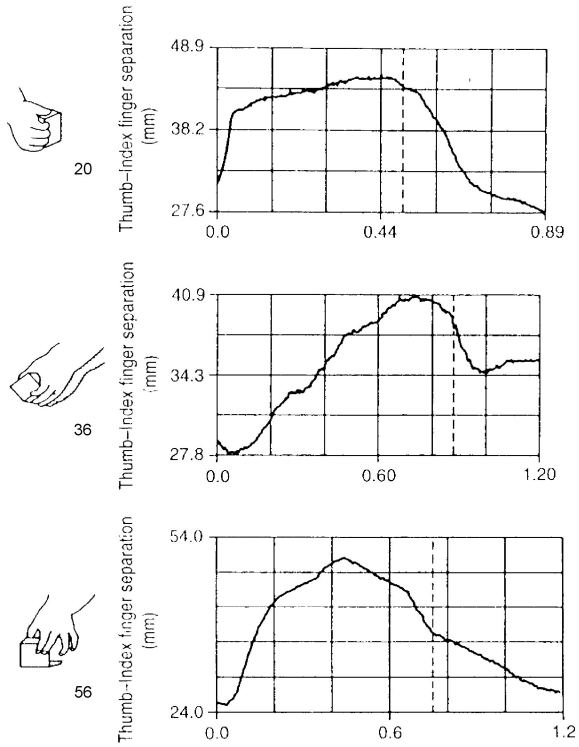


Fig. 3.11 The opening of the hand during reaches for objects. It can be observed that the separation between thumb and index finger increases during the reach and decrease close to the encounter of the object in all three examples (dashed line). Only at the oldest age is the opening and closing of the hand integrated into one continuous movement (from [163]).

An object is optimally grasped over an opposition space that goes through the centre of mass of the object. To investigate infants' tendency to grasp objects in this way, Barrett & Needham [26] presented relatively large symmetrical and asymmetrical objects to 11- and 13-month-old infants. To be able to grasp these objects, infants had to use both hands. The point of contact of each hand was measured and how far the two hands were from the centre of mass of the object. It was found that at first contact, all infants grasped the asymmetrical object further from its centre of mass than the symmetrical object. In addition, results showed that the older infants were better able to correct for less stable hand placements (that is closer to the centre of the object than the centre of mass), to maintain control of the objects without dropping them.

3.3.3.3 Bimanual Coordination

There is no consensus for the definition of what constitutes a bimanual reach. According to Corbetta and Thelen [66], it is enough that both hands move in the approximate direction of the object to constitute a bimanual reach. Other studies require that both hands end up at the object [26]. Another problem has to do with how close in time the two limbs approach the object. If one hand approaches the object a second or more after the first one, it is generally agreed that the reaches should be counted as two separate ones. If the time difference is less, however, the question arises when the reaches with the two hands merged into one bimanual reach. The question is also whether the two hands have to do the same thing or can do complimentary things. An action approach defines a bimanual action as one where both hands serve the same goal. Except when the object is too large to be grasped by a single hand, there is not a certain limit of object size when both hands are needed. Figure 3.12 (a) shows an infant who grasps a large ball with both hands and feet. Fagard et al. [87] found that when grasping and manipulating objects are more task dependent and can be related to object shape and orientation or to the intended action: for instance, banging with one hand and lifting with two. Both hands are more often engaged when the child is reaching for large objects, slippery objects, and moving objects. In some tasks, the object that is being grasped is moved between the hands, in others, one hand assumes support while the other manipulates the object.

While much effort has been devoted to how infants approach and grasp objects, very few studies have focused on the manipulation of objects after they have been grasped. Even when only one hand is used for grasping objects, two hands are most often used for manipulating them. The two hands are also engaged when the subject performs complementary actions like squeezing, tearing, and pulling. Finally, the two hands are engaged when the child performs an action involving two objects like banging one object against another. The object may then be transported from

(a) (b)

Fig. 3.12 (a) A 6-month old child who controls the position of a ball with both hands and feet. (b) An 8-month-old child who manipulates a sheet of paper with both hands.

one hand to the other and back again several times while it is being rotated. It is obvious that the function of such manipulations is to inspect the object from many different angles. Other manipulations that involve both hands are stretching, tearing and wrinkling papers, crumbling bread, and bending and squeezing elastic objects. Figure 3.12 (b) shows a child who is manipulating a shiny piece of paper with both hands. Infants are engaged in such actions from the time they master reaching for and grasping object at around 4 months of age. One hand is most often used when the object, for instance, is banged against a surface. For infants aged between 6 and 36 months Fagard and Lockman [87] studied the use of one or both hands in different conditions: simple grasping, precision grasping, grasping with bimanual manipulation and object exploration. They found that there was a strong decrease in bimanual grasping between 30-36 and 48 months. Increasing the precision required for grasping decreased the variability of the grasping patterns and increased the frequency of right-handed strategies. In contrast, grasping of objects affording various explorations and subsequent exploratory behaviours were even less clearly lateralized than simple grasping. In an object-exploratory task, bimanual use dominates. Exploration was mostly visual-manual at all ages. For banging, one hand was used while exploration that included mouthing of the object engaged both hands most of the time.

Recent studies indicate that laterality doesn't just mature. It is not very stable during the first year of life but rather dependent on the task performed by the infant [300]. If children grasp objects far to the left, the left hand is predominantly used and when they grasp objects far to the right, the right hand is predominantly used. Laterality also has to do with the roles assumed by the two hands. For instance, when opening a jar, the left hand may hold the jar while the right hand unscrews the lid.

Fagard, Spelke, & von Hofsten [88] investigated hand preference, midline crossing, hand cooperation, and visual-field asymmetry in 6-, 8-, and 10-month-old infants who reached for and grasped a moving object by comparing how performance depended on the direction of motion of the object (from right to left versus left to right). It moved on a large circular trajectory in the horizontal plane. The results show that 6-month-olds reached for the object with the ipsilateral hand (from where the object arrived) and grasped it with the contralateral hand. The grasping, but not the reaching, showed a right-hand bias. In the 8-month-olds, the ipsilateral reaching and contralateral grasping was overshadowed by a strong right-hand bias. Finally, the 10-month-olds both reached and grasped preferentially with their ipsilateral hand or with both hands, especially when the object arrived from the left. These age-related changes in reaching strategies seem to be associated with an increase with hand preference coupled with improved manual skills. They support the hypothesis that laterality is more pronounced in a demanding task. The task is difficult for the 6-month-olds and they have not developed very strong hand preference. It is also difficult for the 8-month-olds, who master the task, and with strong expression of laterality. The mastery of the 10-month-olds is more relaxed with weaker laterality.

The results do not support the hypothesis that maturation of manual skills is associated with stronger tendency to cross the midline. On the contrary, Fagard et al. [88] found that midline crossing was most common in the youngest infants and least common in the oldest ones. The results indicate that, in addition to the need for predicting the path of a moving object, motor constraints due to spatial compatibility, hand preference and bimanual coordination must be taken into account to understand age differences in grasping a moving object.

3.3.3.4 Manipulation

The close connection between vision and manipulation makes it also possible to learn about object affordances by viewing other people manipulating them. This is especially relevant when learning about the functions of tools. Lockman [224] suggested that tool used may be a more continuous development than previously believed and that it is rooted in the perception-action routines that infants employ to gain knowledge about their environments. He suggested that in order to learn more about tool use development, research should focus on the processes by which children detect and relate affordances between objects, coordinate spatial frames of reference, and incorporate early-appearing action patterns into instrumental behaviours.

The development of skills in reaching and manipulation are closely related to the development of such cognitive skills as mental rotation and means-end relationships. When manipulating objects, the subjects need to imagine the goal state of the manipulation and the procedures of how to get there. von Hofsten & Örnkloo [279]

Fig. 3.13 A 14-month-old infant who just has placed a triangular shaped block over the triangular aperture. The fact that the block does not fit the aperture in the present orientation does not seem to concern the infant. He is very happy anyway.

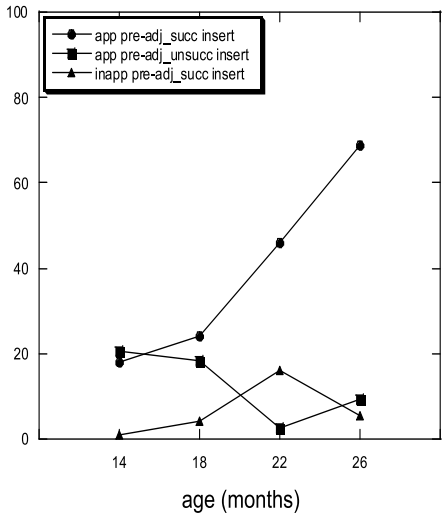


Fig. 3.14 The relationship between pre-adjustments and successful insertions in the block insertion experiment. Circles depict percentages of appropriate pre-adjustments and successful insertions, squares depict appropriate pre-adjustments, and unsuccessful insertions, and triangles unsuccessful pre-adjustments and successful insertions (from [279]).

studied how infants develop their ability to insert blocks into apertures. The task was to insert elongated objects with various cross-sections (circular, square, rectangular, elliptic, and triangular) into apertures in which they fitted snugly. All objects had the same length and the difficulty was manipulated by using different cross sections. The cylinder fitted into the horizontal aperture as long as its longitudinal axis was vertical, while all the other objects also had to be turned in specific ways. The objects were both presented standing up and lying down. It was found that although infants younger than 18 months understood the task of inserting the blocks into the apertures and tried very hard, they had little idea of how to do it. They did not even rise up elongated blocks as shown in Fig. 3.13, but just put them on the aperture and tried to press them in. The 22-month-old children systematically raised up the horizontally placed objects when transporting them to the aperture and the 26-month-olds turned the objects before arriving at the aperture, in such a way that they approximately fit the aperture. This is shown in Fig. 3.14. This achievement is the end point of several important developments that includes motor competence, perception of the spatial relationship between the object and the aperture, mental rotation, anticipation of goal states, and an understanding of means-end relationships. These abilities are not independent of each other in a task like this and cannot be totally separated. Motor competence is expressed in actions and actions rely on spatial perception and anticipations of goal states.

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

3.3.4 Development of Social Abilities

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

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

Social interaction relies primarily on vision, touch, and proprioception. The mouth, face, eyes, and hands are the primary instruments for such actions. There is an important difference between these action systems and those used for negotiating the physical world. The fact that one's own actions affect the behaviour of the person towards whom they are directed creates a much more dynamic situation than when actions are directed towards objects. In addition, anticipating what is going to happen next is less dependent on physical laws as in the object case and more dependent on knowledge of the rules and regularities that govern the other persons' actions that in turn is dependent on one's own social behaviour and social conventions. In order to master social interaction it is therefore crucially necessary to know the conventions of social interaction and perceive the intentions and emotions of the subject with whom one interacts. Intentions and emotions are readily displayed by elaborate and specific movements, gestures, and sounds that become important to perceive and control. Some of these abilities are already present in newborn infants and reflect their preparedness for social interaction. Neonates are very attracted by people, especially to the sounds, movements, and features of the human face [184, 239]. They also engage in social interaction and turn-taking that among other things is expressed in their imitation of facial gestures. Finally, they perceive and communicate emotions such as pain, hunger and disgust through their innate display systems [403]. These innate dispositions give social interaction a flying start and open up a window for the learning of the more intricate regularities of human social behaviour. Parents show a remarkable talent for responding to the

infants' signals and turning them into sophisticated forms of social interaction. Striano & Rochat [360] suggested that this "propensity to express empathy through the echoing of affects and feelings in highly scaffolding ways is part of normal parenting and ... the primary source of intersubjectivity".

Important social information is provided by vision. Primarily, it has to do with perceiving the facial gestures of other people. Such gestures convey information about emotions, intentions, and direction of attention. Perceiving what another person is looking at is an important social skill. It facilitates referential communication. One can comment on objects and immediately be understood by other people, convey information about them, and communicate emotional attitudes towards them [69, 80, 259, 361]. The ability to perceive the gaze direction of others is thus a key component in social communication [262].

Most researchers agree that infants reliably follow gaze from 10-12 months of age [332, 263, 69, 76, 260, 261, 404]. A common method has been to determine the side toward which the infants first turn their gaze (see e.g. [69, 262]). Moore et al. [259], for instance, found that some 9-month-olds, and presumably those with more advanced gaze-following skills, will turn in the direction indicated by a live but static face (left or right). Even 3- to 6-month-old infants have been found to be above chance in following a turning gaze to the correct side [80]. A reasonable conclusion from these studies is that social directional cues can be utilized before 12 months as weak evidence of the probability that some interesting target will be seen to the left or right of the infant. Von Hofsten, Dahlström, & Fredriksson [156] used an eye tracker (TOBII) to study 12-month-old infants' ability to perceive gaze direction in static video images. The images showed a woman who performed attention directing actions by looking and/or pointing towards one of 4 objects positioned in front of her (2 on each side). They found that the infants clearly discriminated the gaze directions to the objects located 10° apart, on the same side of the model. The infants spent more time looking at the attended objects than the unattended ones and they shifted gaze more often from the face of the model to the attended object than to the un-attended objects. In all conditions the infants spent most of the time looking at the model's face. This tendency was especially noticeable in the pointing-only condition and the condition where the model just looked straight ahead.

Humans possess a unique ability to underline their direction of attention by pointing [375]. It is performed with different goals depending on the context. Bates [28] discussed pointing as a way to share the attention in an object (declarative) or to request something (imperative) [28]. The difference is that during imperative pointing infants want to get an object, thereby making the pointing an instrumental gesture, while during declarative pointing they want to share the attention to an interesting object with another person using a socially communicative gesture. This distinction becomes important if we consider that neither of the two types of pointing has been seen naturally in animals, but apes can learn the imperative pointing in captivity [375, 293]. No declarative pointing has been observed in these groups. Still some studies show that the main function of pointing is the declarative one, which cast doubt on the hypothesis that pointing develops from grasping as some authors thought [390]. Pointing has also other functions apart of sharing attention or

requesting an object. Infants as young as 12 months can use it to provide information to adults (the location of an object that the adult was looking for) [223] and there are more possibilities, like asking for information about objects (such as names), indicating a direction, creating imaginary shapes or even to show inferred referents (i.e. pointing to an empty chair to refer to the person who usually sits there) [198]. With this in mind it is not strange that pointing has been studied extensively because of its connection with language. Some studies show that pointing at 12 months predicts speech production rates at 24 months [56] and that the combination of pointing and a word which differs from the object referred to precedes two-word sentences, the first grammatical construction [121]. Also, some researchers indicate how index finger extension is correlated with production of syllabic sounds [231] and that pointing can be the first way to associate the visual object with a sound.

As we get more information about pointing, we are left with many unanswered questions. We know that between 8 and 13 months infants begin pointing to significant objects, but there is a current discussion on why infants start to point. The onset may be innate or start through imitation when infants see other people point. The onset can also be conditioned by the presence of the object they want (imperative pointing) or by the parents' positive reaction and shared attention (declarative pointing) [57]. There is also an active discussion whether the declarative or imperative pointing comes first. Some authors think that infants probably comprehend pointing one month before they perform pointing themselves and others state that infants start pointing before they follow other's pointing [231].

The most important perception-action systems that serves social interaction is speech. Like other action systems, speech has both a perceptual and a productive side. Perception of certain aspects of speech seems to occur in the womb already and newborn infants have been shown to prefer their mother's voice [77]. Because of the lowpass filtering of the human voice in the womb, it is presumably the prosody of the voice rather than any other more detailed property that neonates recognize. There is good evidence that infants are sensitive to prosodic structure and that this sensitivity is present in the newborn [188]. Also the phonemic structure develops early. By 4 months of age, infants seem to be able to distinguish between virtually any pair of stimuli that crosses phoneme boundaries [208].

The research on early development of speech shows that the productive capabilities of speech clearly lag the perceptual ones (see e.g. [247]). Thus, human infants can perceive speech before the can speak or babble. On the other hand, phylogeny has prepared the human child for the task of speaking. The morphology of the human vocal tract has been altered relative to that of other primates in a way that facilitates speech [58]. Babbling is, furthermore, dominated by the cyclical opening and closing of the mandible in a way that is also characteristic of sucking. MacNeilage & Davis [227] argued that many of the articulatory regularities in the sound patterns of babbling and early speech can be attributed to properties of this mandibular cycle. During the second half of the first year of life, infants spend much of their time awake exercising babbling sounds. They also discover the communicative value of speech sounds and use them in their social interactions much before they can articulate specific words. In addition to this, infants start pointing at around 11 months

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

sulcus, while Brodmann's area 6 comprises a mosaic of at least six different areas. According to the nomenclature proposed by Matelli and collaborators at the end of the Eighties [232, 233, 234], areas F2 and F7 form the dorsal sector of area 6, areas F4 and F5 the ventral sectors, and areas F3 (corresponding to Woolsey's supplementary motor area, MII) and F6 lie on the mesial face of the hemispheres, within the interhemispheric scissure. A similar organisation can be also seen in the cortex of the lower parietal lobule and the two mosaics, the frontal one and the parietal one, are bi-directionally connected through a series of parallel circuits.

2. The electrophysiological study of the neurons at the origin of the parietofrontal parallel circuits shows they play a role in converting sensorial information into actions specifically organised for a certain effector (e.g. the arm, the hand, the head or the eyes) and not according to a specific sensory modality (visual, somestesic, etc.). Consequently, Penfield's and Woolsey's original concept, according to which just two somatotopically organised motor areas exist in the frontal lobe, is now insufficient to explain this new experimental evidence. On the contrary, every effector is represented in the cortex several times, and this plurality of representations is justified by the plurality of the sensory information (visual, proprioceptive, tactile). For example, despite the fact that the electrical microstimulation of specific regions of area F1 (primary motor), area F2 (dorsal premotor) and area F4 (ventral premotor) always generates arm movements, area F2 is reached by mainly proprioceptive information, whereas area F4 is reached by visual and tactile information.
3. Apart from their peculiar motor activity, various neurons located in the agranular frontal cortex also discharge during sensory stimulation. This sensorimotor coupling is often effector-specific: switching on a light stimulus within a specific region of the visual field excites motor neurons in FEF, deep proprioceptive and tactile stimulation of the hand generates neuronal responses in the primary motor cortex; tactile surface stimulation of the face, arms and trunk activates the neurons in area F4; passive mobilization of the arm excites neurons in area F2; passive stimulation of the fingers correlates with activity of neurons in area F5. Some motor neurons, especially in the ventral part of the premotor cortex, are even activated by visual stimuli.

This latter observation, in particular, raises important questions about the nature of the activity of these "sensorimotor neurons". The fact that their output signal is the same during motor execution and during sensory stimulation suggests an apparent functional ambiguity. In practice, the stimulus-response association characterizing the sensorimotor neurons of the frontal cortex might provide the movement with the goal. This would convert the representation of the movements, as originally conceived by the classical physiologist, into representation of actions. By this we do not mean that there are not exclusively motor neurons as well. However, they live together with the sensorimotor ones within the same cortical areas. On the contrary, it is not yet known if there is a cortical area that contains just motor neurons.

4.2 The System for Grasping

In the recent years several neurophysiological studies have given us a good overall view of how reaching and grasping actions are planned and executed by monkey brain. The premotor area used to program and control prehension is area F5, located in the most anterior part of the ventral premotor cortex. The intracortical microstimulation of area F5 and the recording of individual neurons during active motion indicate that a representation of hand and mouth movements exists in this area [276, 311]. Area F5 is strongly connected with the inferior parietal lobule, namely with areas AIP (particularly the subsector F5c) and PF/PFG (F5a).

Most neurons of area F5 become active during finalized movements, such as grasping, manipulating, tearing and holding [311]), but not for hand or finger movements produced with similar muscular patterns but used for other purposes (e.g. scratching, pushing away from the body, etc.). Moreover, many neurons in F5 activate during movements having the same goal but performed with different effectors (e.g. grasping an object with the right hand, the left hand or the mouth). In this case, clearly, there is not much point in attempting to describe neuron activation in terms of elementary movements.

Neurons mostly represented in area F5 are those specialized in "grasping". Typically, these neurons begin to discharge before contact is made between the hand and the object to grasp and many of them discharge in relation to a particular type of prehension. For example, the grasping of a sphere (which requires the opposition of all the fingers) is codified by neurons different from those that codify the grasping of a cylinder (which requires the opposition of all the fingers except for the thumb).

Considered in overall terms, the functional properties of F5 neurons suggest that this cortical area stores a set of motor schemes [10], or (as previously suggested [317]) a "vocabulary" of motor acts. The "words" comprising this vocabulary are populations of neurons: some indicate the general category of an action (hold, grasp, tear, manipulate); others specify the appropriate methods for adapting the hand to the object to hold (e.g. specific neurons for precision grasping or for whole hand prehension); others, lastly, are dedicated to the temporal segmentation of the action (opening the hand, closing the fingers, etc.). What differentiates the area F5 from the primary motor cortex (M1) is that the motor "words" are finalized actions or portions of finalized actions, while M1 stores movements regardless of the context in which they are performed. Compared with F5, area M1 could be defined as a "dictionary of movements".

The above motor properties are common to all the F5 neurons. However, if one examines the ones that also respond to visual stimuli, it is evident that two categories of visual motor neurons exist in F5. The neurons in the first category discharge when the monkey observes graspable objects ("canonical" F5 neurons [313]). The neurons in the second category discharge when the monkey observes another individual performing an action similar to that they motorically code [283, 109, 315]. For these peculiar properties of "action resonance", the neurons belonging to the second

category have been defined as “mirror” neurons [109]. The finding that an elevated percentage of the tested neurons respond to the visual presentation of graspable three-dimensional objects, makes it possible to set an interpretative code of the manipulable environment already during simple observation. Approximately two thirds of these visuomotor neurons are selective for one or more objects [267]. The second class of F5 visuomotor neurons comprises the mirror neurons that discharge when the monkey observes another individual performing actions involving the hand or the mouth. They will be not further described here.

However, both canonical and mirror neurons show an interesting congruence between the action motorically coded by a given neuron and the object/action which, when observed, can evoke the discharge of the same neuron. How can these findings be explained? Obviously, the visual responses correlated to an object (or to an action performed by another individual) do not depend on motor preparation: Why do the canonical neurons discharge during the visual presentation of objects even if no request to grasp them has been made? Why do mirror neurons respond to the observation of another monkey grasping a raisin, even if this means that the raisin has become definitely unavailable? The most plausible interpretation of the visual discharge of the canonical neurons is that — at least in adult individuals — there is a close connection between the most common three-dimensional visual stimuli and the actions required to interact with them. Consequently, whenever an object that can be grasped is presented, the corresponding F5 neurons are activated and a motor program is evoked. It is interesting to note that a similar phenomenon was also recently observed in humans during a brain imaging study: the presentation of tools or objects that can be grasped activates the premotor cortex even when no motor response is requested [210].

Therefore, motor activation in relation to an object seems to represent (in monkeys and humans) a potential action, an “idea” of how to act. In certain circumstances, it guides the execution of the movement, in others it is an unexecuted representation that can also be used as semantic knowledge. In other words, according to this interpretation, the discharge of the visuomotor neurons can simply imply that a particular action is internally represented, regardless of the future use that the brain can make of it: actions can “come to mind” when we observe objects that we can grasp, when we observe other individuals acting, when we think about acting and — obviously — when we decide to act. Only in the latter circumstance are potential actions transformed into real ones.

The possibility that the nervous system internally represents an action in the absence of motor contingencies opens up exciting new prospects also for the interpretation of some perceptual mechanisms. The visuomotor responses of the F5 canonical neurons could, e.g., play also a role in the semantic categorization of objects [86]. Though semantic analysis of objects probably involves temporal areas in the so-called “ventral stream” [379], it is also plausible that a complete semantic knowledge of objects must take into consideration the information concerning the method used to act on them.

4.3 The Distributed Perception of Space

Since the early Eighties, the dominant view on the cortical processing of visual information has been the 'what' and 'where' theory, as formalized by Ungerleider and Mishkin [379]. According to these authors, the ventral stream has its main role in object recognition, while the dorsal stream analyzes object's spatial location. This point of view was in accordance with the classical notion of the parietal cortex as the site for unitary space perception, used for all purposes: for walking, for reaching objects, for describing a scene verbally. Lesions of this lobe and, in particular, of the inferior parietal lobule produce a series of spatial deficit ranging from space distortions to spatial neglect.

Since 1991, Milner and Goodale have argued against this theory, emphasizing the pragmatic role of the dorsal stream. This point of view, primarily triggered by clinical data, has been subsequently substantiated by neurophysiological evidence. The posterior parietal cortex, as pointed out also by Milner and Goodale [253], is constituted by a mosaic of independent areas. If one of these areas were the hypothetical space master centre, it should be also the centre of a series of convergent and divergent connections. It should receive inputs from the occipital lobe and distribute its output to a variety of other brain centres: oculomotor centres for looking at the objects, areas controlling walking for navigating in the environment, and so on. The evidence is exactly the opposite. The connections of parietal lobe with the frontal lobe as well as with subcortical centres are remarkably segregated ([4, 59, 232, 287]). For example, the connections of parietal area LIP (lateral intraparietal) are exclusively or almost exclusively with Brodmann area 8 (frontal eye fields, FEF). Both these areas are related to oculomotion. Area LIP, in contrast, does not send any input to areas related to arm movements. Thus there is no evidence of a unique supramodal "space area" within the posterior parietal cortex. Space perception appears to derive from the joint activity of a series of sensorimotor fronto-parietal circuits, each of which, according to its own motor purposes, encodes the spatial location of an object and transforms it into a potential action (see [316, 314]).

The idea of a motor role for the posterior parietal cortex is by no means new. Since the pioneering studies carried out by Hyvarinen, Mountcastle and their coworkers [176, 265] it is well known that different sectors of the posterior parietal cortex are involved in the control of arm, hand and eye movements. However, the 'motor' role was somehow underestimated in light of a purely 'spatial' characterization of the visual information reaching these sectors of the parietal cortex.

Milner and Goodale [253] make two major points: 1) The dorsal stream processes visual information for motor purposes; 2) Action and perception are two completely separate domains, the latter being an exclusive property of the ventral stream. While a consistent set of neurophysiological data confirm the 'pragmatic' role of the visual information processed in the dorsal stream, and thus corroborates the theoretical views of Milner and Goodale [253], a series of neurophysiological, neuropsychological and brain imaging studies suggests that the dichotomy proposed by Milner and Goodale between action and perception is probably too rigid.

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

The LIP-FEF circuit contains three main classes of neurons: neurons responding to visual stimuli (visual neurons), neurons firing in association with eye movements (motor neurons), and neurons with both visual- and movement-related activity (visuomotor neurons) [5, 53, 120]. Neurons responsive to visual stimuli respond vigorously to stationary light stimuli. Their receptive fields (RFs) are usually large. Movement-related neurons fire in relation to ocular saccades, most of them discharging before the saccade onset. Visuomotor neurons have both visual- and saccade-related activity. Visual RFs and 'motor' fields are in register, that is, the visual RF corresponds to the end-point of the effective saccade. Visual responses in both LIP and FEF neurons are coded in retinotopic coordinates [5, 120]. In other words, their RFs have a specific position on the retina in reference to the fovea. When the eyes move, the RF also moves. Most LIP neurons have, however, an important property. The intensity of their discharge is modulated by the position of the eye in the orbit (orbital effect). Now, if the position of the RF on the retina and the position of the eye in the orbit are both known, one can reconstruct the position of the stimulus in spatial (craniocentric) coordinates. Thus, although the firing of a neuron does not specify by itself the position of the triggering stimulus in space, the spatial location of stimulus can be derived from the discharge intensity of different neurons [52]. Specifically, 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.

As in the LIP-FEF circuit, neurons in the VIP-F4 circuit can be subdivided into three main classes: sensory neurons, movement-related neurons, and sensorimotor neurons. The majority of them belong to the last category. Movement-related neurons and sensorimotor neurons are activated by head movements, face movements, or arm movements. Sensory and sensorimotor neurons respond to tactile or to tactile and visual stimuli. The visual RFs of these neurons are anchored to the tactile ones regardless of eye position [115, 114]. F4 neurons fire tonically at the presentation of stationary three-dimensional objects within monkey peripersonal space. A very intriguing finding is that some of these tonically discharging neurons continue to fire when, unknown to the monkey, the stimulus previously presented has been withdrawn, and the monkey 'believes' that it is still near its body. Space representation in the premotor cortex can be generated, therefore, not only as a consequence of an external stimulation but also internally on the basis of previous experience [129].

If we now compare the properties of the VIP-F4 circuit with those of the LIP-FEF circuit, we find a common aspect and some important differences. The common aspect is that coding of space is not devoted to a multiplicity of purposes but is specifically directed to a particular motor goal: eye movements in the case of

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

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

4.4 Perception Depends on Action

As we noted already, Milner and Goodale [253] maintain that in the primate's visual system there is a sharp distinction between the role played by the dorsal and the ventral stream of visual processing with the dorsal stream being mainly involved in the on-line control of actions and with the ventral stream being the exclusive source of information for perception and semantics. However, several lines of evidence seem to point to an important involvement of the motor system in supporting processes traditionally considered to be 'high level' or cognitive, such as action understanding, mental imagery of actions, and perceiving and discriminating objects. A first example is provided by the discovery of a population of neurons in the monkey ventral premotor cortex (mirror neurons) that discharge both when the monkey performs a grasping action and when it observes the same action performed by other individuals [109]. Mirror neurons provide the neurophysiological basis for the capacity of primates to recognize different actions made by other individuals: the same motor pattern which characterizes the observed action is evoked in the observer as when activates its own motor repertoire. This matching mechanism, which can be framed within the motor theories of perception, offers the great advantage of using a repertoire of coded actions in two ways at the same time: at the output side

to act, and at the input side, to analyse the visual percept. This matching system has also been demonstrated in humans. Transcranial Magnetic Stimulation (TMS) of the motor cortex of subjects observing hand actions made by the experimenter determined an enhancement of motor evoked potentials (MEPs) in the same muscular groups that were used by the experimenter in executing those actions [85]. This means that when we observe an action we utilize, as monkeys do, the repertoire of motor representations used to produce the same action.

Another example of the involvement of the dorsal stream in cognitive functions is motor imagery. Imagining a grasping action is a cognitive task that requires a conscious, detailed representation of the movement. Several PET studies have shown that during motor imagery of grasping actions premotor and inferior parietal areas are strongly activated [78, 122]. Furthermore, Parsons et al. [281] demonstrated in a PET study that motor imagery used for visual hand shape discrimination activates premotor and posterior parietal cortex.

Further evidence supporting the notion of the involvement of the dorsal stream in cognitive tasks is provided by an elegant neuropsychological study by Sirigu et al. [346]. Patients with lesions restricted to the posterior parietal cortex were selectively impaired at predicting through mental imagery the time necessary to perform differentiated finger movements.

The role played by handedness in performing cognitive tasks is another example of the involvement of motor processes in perceptual functions [357] showed that right- and left-handed normal subjects used an internal simulation of the movement of their dominant hand in order to discriminate between observed screwing and unscrewing screwdrivers. In another series of experiments [112, 113], normal subjects were required to judge handedness of pictures of hands and fingers assuming different postures. The results showed that the presentation of postures that hand and fingers commonly assume at rest, or when interacting with objects, facilitated the responses with respect to the presentation of less usual hand-finger postures, even when the latter were richer in visual cues useful for handedness recognition. Once again procedural motor knowledge was employed to solve a cognitive task.

Taken together, all these results seem to contradict a sharp distinction between an ‘acting brain’ and a ‘knowing brain’.

4.5 Action and Selective Attention

Among the processes traditionally considered to be ‘high level’ or cognitive, selective attention is one of the most important. It refers to the capability of selecting a particular stimulus according to its physical properties, way of presentation, or previous contingencies and instructions. After selection, the stimulus is processed and, if convenient for the individual, acted on. According to the scenario for space representation described above, a problem is how the different sectors of space representation can increase their efficiency in processing visual stimuli in order to select some of them and discard others.

The traditional view is that selective attention is controlled by a supramodal system 'anatomically separate from the data processing systems' ([295], p. 26). Like the sensory and motor systems, this 'attention system' performs operations on specific inputs. It interacts with other centres of the brain but maintains its own identity [295]. However, on the basis of data obtained from brain imaging experiments [67, 68, 296], it has been suggested that the attention system is not unitary but consists of at least two independent systems: a posterior one subserving spatial attention and an anterior one devoted to attention recruitment and control of brain areas involved in complex cognitive tasks [294].

Another view of selective attention is that it derives from mechanisms that are intrinsic to the circuits underlying perception and action. Attention is modular and there is no need to postulate control mechanisms anatomically separate from the sensorimotor circuits. This account of selective attention was originally formulated for spatial attention (premotor theory of attention; [310, 318]) and it is deeply rooted in the idea that space is coded in a series of parieto-frontal circuits working in parallel and that the coordinate frame in which space is coded depends on the motor requirements of the effectors that a given circuit controls (see [319]). Given this strict link between space coding and action programming, the premotor theory of attention postulates that spatial attention is a consequence of an activation of those cortical circuits and subcortical centres that are involved in the transformation of spatial information into action. Its main assumption is that the motor programs for acting in space, once prepared, are not immediately executed. The condition in which action is ready but its execution is delayed corresponds to what is introspectively called spatial attention. In this condition, two events occur: (a) there is an increase in motor readiness to act in the direction of the space region toward which a motor program was prepared, and (b) the processing of stimuli coming from that same space sector is facilitated. There is no need, therefore, to postulate an independent control system. Attention derives from the mechanisms that generate action. Although, in principle, all circuits responsible for spatially directed action can influence spatial attention, there is no doubt that in humans the central role in spatial attention is played by the circuits that code space for programming eye movements. Experiments in which the relations between attention and eye movements were either indirectly or directly tested showed that the two mechanisms interact: Any time attention is directed to a target, an oculomotor program toward that target is prepared. Particularly significant in this respect are experiments in which the relations between attention and eye movements were directly tested [342, 341]. Sheliga and co-workers instructed normal participants to pay attention to a given spatial location and to perform a predetermined vertical or horizontal ocular saccade at the presentation of the imperative stimulus. Results showed that the trajectory of ocular saccades in response to visual or acoustic imperative stimuli deviates according to the location of attention. The deviation increased as the attentional task became more difficult. Note that if spatial attention were independent of oculomotor programming, ocular saccades

should not be influenced by location of attention. In a recent experiment, the role of oculomotion in orienting of attention was investigated by dissociating perceptual from motor capabilities [71]. If a causal relationship links oculomotion and orienting of attention, any constraint limiting eye movements should abolish, or at least reduce, attentional benefits in the region of the spatial field barely reachable by the eye. On the contrary, if attention is a purely cognitive process, then no effects are expected to arise from oculomotor constraints. Subjects were submitted to a spatial attention orienting task, performing it in monocular vision and having the head rotated in such a way that the eye was kept at an extreme position in the orbit. This position limited the execution of a saccade toward the temporal hemifield, whereas it allowed saccadic execution toward the nasal hemifield. Results showed that orienting of attention was normal in the nasal but not in the temporal hemifield, indicating that eyes and attention show a common limit stop. In other words, oculomotor constraints on saccadic actions induce similar constraints on spatial attention.

In primates, eye movements are certainly the most important mechanism for selecting stimuli. However, there are also circumstances (e.g., stimuli presented very close to the face) in which eye movements are not crucial for selecting stimuli in space. In these circumstances, spatial attention should depend on circuits other than those related to eye movements. Probably the best documented evidence in favor of spatial attention not related to eye movements is that deriving from experiments conducted by Tipper et al. [374]. They studied, in normal participants, the effect of an irrelevant stimulus located inside or outside of the arm trajectory necessary to execute a pointing response. The results showed that an interference effect was present only when the distractor was located inside the trajectory of the arm. Control experiments suggested that the effect was not due to a purely visual representation of the stimuli or to spatial attention related to eye movements. Rather, the organization of the arm-hand movement determined a change in the attentional relevance of stimuli close to the hand or far from it. In other words, if something is within reach, attention to it is enhanced; if not, attention is diminished.

In the frame of premotor theory of attention, Craighero and colleagues [70] assumed that allocation of attention to a graspable object is a consequence of preparing a grasping movement to that same object. The authors predicted that, when a specific grasping movement was activated, there would be both an increase in the motor readiness to execute that movement and a facilitation in visual processing of graspable objects the intrinsic properties of which are congruent with the prepared grasping. In the experiment, normal subjects were required to grasp a bar after the presentation of a visual stimulus whose orientation was either congruent or incongruent with that of the bar. The results supported the hypothesis. The detection of a visual object was facilitated by the preparation of a grasping movement congruent with the object's intrinsic properties. This finding strongly suggests that the premotor theory of attention is not limited to orienting attention to a spatial location but can be generalized to the orienting of attention to any object that can be acted on.

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

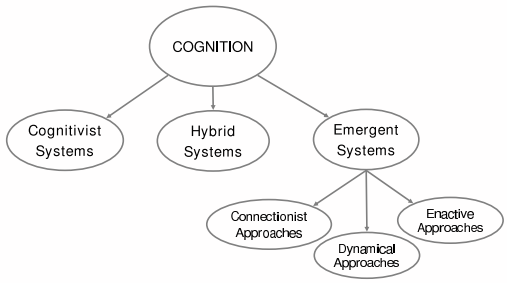


Fig. 5.1 The cognitivist, emergent, and hybrid paradigms of cognition

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].

Emergent systems exploit processes of self-organization, self-production, self-maintenance, and development, through the concurrent interaction of a network of distributed interacting components.

Representational Framework

Cognitivist systems use patterns of symbol tokens that refer to events in the external world. These are typically the descriptive[3] product of a human designer and are usually, but not necessarily, punctate rather than distributed.

Emergent systems representations are global system states encoded in the dynamic organization of the system's distributed network of components.

Semantic Grounding

Cognitivist systems symbolic representations are grounded through percept-symbol identification by either the designer or by learned association. These representations are accessible to direct human interpretation.

Emergent systems ground representations by autonomy-preserving anticipatory and adaptive skill construction. These representations only have meaning insofar as they contribute to the continued viability of the system and are inaccessible to direct human interpretation.

Temporal Constraints

Cognitivist systems are atemporal and are not necessarily entrained by the events in the external world.

Emergent systems are entrained and operate synchronously in real-time with events in its environment.

Inter-agent Epistemology

For cognitivist systems an absolute shared epistemology between agents is guaranteed by virtue of their positivist view of reality; that is, each agent is embedded in an environment, the structure and semantics of which are independent of the system's cognition.

The epistemology of emergent systems is the subjective agent-specific outcome of a history of shared consensual experiences among phylogentically-compatible agents.

Embodiment

Cognitivist systems do not need to be embodied, in principle, by virtue of their roots in functionalism (which holds that cognition is independent of the physical platform in which it is implemented [105]).

Emergent systems are intrinsically embodied and the physical instantiation plays a direct constitutive role in the cognitive process [383, 204, 111].

Perception

In cognitivist systems, perception provides an interface between the absolute external world and the symbolic representation of that world. The role of perception os to abstract faithful spatio-temporal representations of the external world from sensory data.

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

In emergent systems, perception is an agent-specific interpretation of perturbations of the system by the environment.

Action

In cognitivist systems, actions are causal consequences of symbolic processing of internal representations.

In emergent systems, actions are perturbations of the environment by the system, typically to maintain the viability of the system.

Anticipation

In cognitivist systems, anticipation typically takes the form of planning using some form of procedural or probabilistic reasoning with some a priori model.

Anticipation in the emergent paradigm requires the system to visit a number of states in its self-constructed perception-action state space without commiting to the associated actions.

Adaptation

For cognitivism, adaptation ususally implies the acquisition of new knowledge.

In emergent systems, adaptation entails a structural alteration or re-organization to effect a new set of dynamics. Adaptation can take the from of either learning or development.

Motivation

In cognitivist systems, motives provide the criteria which are used to select the goal to adopt and the associated actions.

In emergent systems, motives encapsulate the implicit value system that modulate the system dynamics of self-maintenance and self-development, impinging on perception (through attention), action (through action selection), and adaptation (through the mechanisms that govern change), such as enlarging the space of viable interaction.

Autonomy

Autonomy[4] The cognitivist paradigm does not necessarily entail autonomy. The emergent paradigm does since cognition is the process whereby an autonomous system becomes viable and effective through a spectrum of homeostatic processes.

Cognition

In the cognitivist paradigm, cognition is the rational process by which goals are achieved by reasoning with symbolic knowledge representations of the world in which the agent operates.

In the emergent paradigm, cognition is the dynamic process by which the system acts to maintain its identity and organizational coherence in the face of environmental perturbation. Cognition entails system development to improve its anticipatory capabilities and extend it space of autonomy-preserving actions.

[4] There are many possible definitions of autonomy, ranging from the ability of a system to contribute to its own persistence [40] through to the self-maintaining organizational characteristic of living creatures — dissipative far-from equilibrium systems — that enables them to use their own capacities to manage their interactions with the world, and with themselves, in order to remain viable [61].

Table 5.1 A comparison of cognitivist and emergent paradigms of cognition; refer to the text for a full explanation (adapted from [387] and extended)

The Cognitivist Paradigm vs. **the Emergent Paradigm**		
Characteristic	**Cognitivist**	**Emergent**
Computational Operation	Syntactic manipulation of symbols	Concurrent self-organization of a network
Representational Framework	Patterns of symbol tokens	Global system states
Semantic Grounding	Percept-symbol association	Skill construction
Temporal Constraints	Atemporal	Synchronous real-time entrainment
Inter-agent epistemology	Agent-independent	Agent-dependent
Embodiment	No role implied: functionalist	Direct constitutive role: non-functionalist
Perception	Abstract symbolic representations	Perturbation by the environment
Action	Causal consequence of symbol manipulation	Perturbation by the system
Anticipation	Procedural or probabilistic reasoning	Traverse of perception-action state space
Adaptation	Learn new knowledge	Develop new dynamics
Motivation	Criteria for goal selection	Increase space of interaction
Autonomy	Not entailed	Cognition entails autonomy
Cognition	Rational goal-achievement	Self-maintenance and self-development
Philosophical Foundation	Positivism	Phenomenology

Philosophical Foundations

The cognitivist paradigm is grounded in positivism [105].

The emergent paradigm is grounded in phenomenology [108, 386].

5.1.1 The Cognitivist Paradigm

Cognitivism holds that cognition is achieved by computations performed on internal symbolic knowledge representations whereby information about the world is abstracted by perception, and represented using some appropriate symbolic data-structure, reasoned about, and then used to plan and act in the world. The approach has also been labelled by many as the *information processing* (or symbol manipulation) approach to cognition [230, 272, 136, 291, 197, 382, 370, 195] In most cognitivist approaches concerned with the creation of artificial cognitive systems, the symbolic representations are the descriptive product of a human designer. This is significant because it means that they can be directly accessed and understood or interpreted by humans and that semantic knowledge can be embedded directly into and extracted directly from the system. Cognitivism makes the positivist 'the world we perceive is isomorphic with our perceptions of it as a geometric environment' [343]. In cognitivism, the goal of cognition is to reason symbolically about these representations in order to effect the required adaptive, anticipatory, goal-directed,

behaviour. Typically, this approach to cognition will deploy an arsenal of techniques including machine learning, probabilistic modelling, and other techniques in an attempt to deal with the inherently uncertain, time-varying, and incomplete nature of the sensory data that is being used to drive this representational framework. However, this doesn't alter the fact that the representational structure is still predicated on the descriptions of the designers. In cognitivist systems, the instantiation of the computational model of cognition is inconsequential: any physical platform that supports the performance of the required symbolic computations will suffice. This divorce of operation from instantiation is known as functionalism.

5.1.2 The Emergent Paradigm

In the emergent paradigm, cognition is the process whereby an autonomous system becomes viable and effective in its environment. It does so through a process of self-organization through which the system is continually maintaining its operational identity through moderation of mutual system-environment interaction and co-determination [237]. Co-determination implies that the cognitive process determines what is real or meaningful for the agent: the agent constructs its reality (its world and the meaning of its perceptions and actions) as a result of its operation in that world. Thus, cognitive behaviour is sometimes defined as the automatic induction of an ontology: such an ontology will be inherently specific to the embodiment and dependent on the system's history of interactions, i.e., its experiences. Thus, for emergent approaches, perception is concerned with the acquisition of sensory data in order to enable effective action [237] and is dependent on the richness of the action interface [124]. Sandini et al. have argued that cognition is also the complement of perception [329]. Perception deals with the immediate and cognition deals with longer timeframes. Thus cognition reflects the mechanism by which an agent compensates for the immediate nature of perception and can therefore adapt to and anticipate environmental action that occurs over much longer timescales. That is, cognition is intrinsically linked with the ability of an agent to act prospectively: to operate in the future and deal with what might be, not just what is. In contrast to the cognitivist approach, many emergent approaches assert that the primary model for cognitive learning is anticipative skill construction rather than knowledge acquisition and that processes which both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems [61]. While cognitivism entails a self-contained abstract model that is disembodied in principle, the physical instantiation of the systems plays no part in the model of cognition [383, 384]. In contrast, emergent systems are intrinsically embodied and the physical instantiation plays a pivotal role in cognition. They are neither functionalist nor positivist.

5.1.3 The Hybrid Paradigm

Considerable effort has also gone into developing approaches which combine aspects of the emergent systems and cognitivist systems [124, 125, 126]. Typically,

hybrid systems exploit symbolic knowledge to represent the agent's world and logical rule-based systems to reason about this knowledge in order to achieve goals and select actions while at the same time using emergent models of perception and action to explore the world and build these representations. Hybrid systems still use representations and representational invariances but these representations are constructed by the system itself as it interacts with and explores the world rather than through a priori specification or programming. Thus, objects are represented as 'invariant combinations of percepts and responses where the invariances (which are not restricted to geometric properties) need to be learned through interaction rather than specified or programmed a priori' [124]. Thus, just like emergent systems, the agent's ability to understand the external world is dependent on its ability to flexibly interact with it and interaction is an organizing mechanism that drives a coherence of association between perception and action. Thus, hybrid systems are in many ways consistent with emergent systems while still exploiting programmer-centred representational frameworks (for example, see [277]).

5.1.4 Relative Strengths

The foregoing paradigms have their own strengths and weaknesses, their proponents and critics, and they stand at different stages of scientific maturity. The arguments in favour of emergent systems are compelling but the current capabilities of cognitivist systems are actually more advanced.

Several authors have provided detailed critiques of the various approaches. These include, for example, Clark [63], Christensen and Hooker [62], and Crutchfield [72].[5]

Christiansen and Hooker argued [62] that cognitivist systems suffer from three problems: the symbol grounding problem, the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new data),[6] and the combinatorial problem. These problems are one of the reasons why cognitivist models have difficulties in creating systems that exhibit robust sensorimotor interactions in complex, noisy, dynamic environments. They also have difficulties modelling the higher-order cognitive abilities such as generalization, creativity, and learning [62]. According to the Christensen and Hooker, and as we have remarked on several occasions, cognitivist systems are poor at functioning effectively outside narrow, well-defined problem domains.

Emergent systems should in theory be much less brittle because they emerge — and develop — through mutual specification and co-determination with the environment, but our ability to build artificial cognitive systems based on these principles is actually very limited at present. To date, dynamical systems theory has provided

[5] The following is abstracted from [387].

[6] In the cognitivist paradigm, the frame problem has been expressed in slightly different but essentially equivalent terms: how can one build a program capable of inferring the effects of an action without reasoning explicitly about all its perhaps very many non-effects? [338].

more of a general modelling framework rather than a model of cognition [62] and has so far been employed more as an analysis tool than as a tool for the design and synthesis of cognitive systems [251, 62]. The extent to which this will change, and the speed with which it will do so, is uncertain. Hybrid approaches appear, to some at least, to offer the best of both worlds: the adaptability of emergent systems (because they populate their representational frameworks through learning and experience) but the advanced starting point of cognitivist systems (because the representational invariances and representational frameworks don't have to be learned but are designed in). However, it is unclear how well one can combine what are ultimately highly antagonistic underlying philosophies. Opinion is divided, with arguments both for (e.g. [63, 72, 123]) and against (e.g. [62]).

Clark suggests that one way forward is the development of a form of 'dynamic computationalism' in which dynamical elements form part of an information-processing system [63]. This idea is echoed by Crutchfield [72] who, whilst agreeing that dynamics are certainly involved in cognition, argues that dynamics per se are "not a substitute for information processing and computation in cognitive processes" but neither are the two approaches incompatible. He holds that a synthesis of the two can be developed to provide an approach that does allow dynamical state space structures to support computation. He proposes 'computational mechanics' as the way to tackle this synthesis of dynamics and computation. However, this development requires that dynamics itself needs to be extended significantly from one which is deterministic, low-dimensional, and time asymptotic, to one which is stochastic, distributed and high dimensional, and reacts over transient rather than asymptotic time scales. In addition, the identification of computation with digital or discrete computation has to be relaxed to allow for other interpretations of what it is to compute.

It might be opportune to remark at this point on the dichotomy between cognitivist and emergent systems. As we have seen, there are some fundamental differences these two general paradigms — the principled disembodiment of physical symbol systems vs. the mandatory embodiment of emergent developmental systems [384], and the manner in which cognitivist systems often preempt development by embedding externally-derived domain knowledge and processing structures, for example — but the gap between the two shows some signs of narrowing. This is mainly due (i) to a fairly recent movement on the part of proponents of the cognitivist paradigm to assert the fundamentally important role played by action and perception in the realization of a cognitive system; (ii) to the move away from the view that internal symbolic representations are the only valid form of respresentation [63]; and (iii) to the weakening of the dependence on embedded a priori knowledge and the attendant increased reliance on machine learning and statistical frameworks both for tuning system parameters and the acquisition of new knowledge both for the representation of objects and the formation of new representations. However, cognitivist systems still have some way to go to address the issue of true ontogenetic development with all that it entails for autonomy, embodiment, architecture plasticity, and system-centred construction of knowledge mediated by exploratory and social motivations and innate value systems.

5.2 Cognitive Architectures

Any cognitive system is inevitably going to be complex. Nonetheless, it is also the case that it will exhibit some degree of structure. This structure is often encapsulated in what is known as a cognitive architecture.

Although used freely by proponents of the cognitivist, emergent, and hybrid approaches to cognitive systems, the term *cognitive architecture* originated with the seminal cognitivist work of Newell et al. [270, 271, 326]. Consequently, the term has a very specific meaning in this paradigm where cognitive architectures represent attempts to create unified theories of cognition [55, 271, 7], i.e. theories that cover a broad range of cognitive issues, such as attention, memory, problem solving, decision making, learning, from several aspects including psychology, neuroscience, and computer science. Newell's Soar architecture [211, 326, 220, 222], Anderson's ACT-R architecture [6, 7], Sun's CLARION architecture [362, 363], and Minsky's *Society of Mind* [254] are all candidate unified theories of cognition.

Since unified theories of cognition are concerned with the computational understanding of human cognition, cognitivist cognitive architectures are also concerned with human cognition [271, 364]. There is an argument that that the term cognitive architecture should be reserved for systems that model human cognition, suggesting that the term "agent architecture" as a better term to refer to general intelligent behaviour, including both human and machine cognition [407]. However, it has become common-place to use the term cognitive architecture in this more general sense, both in the cognitivist and emergent paradigms, so we will use it in this generic sense throughout the book on the understanding that a cognitive architecture may entail different requirements and characteristics, depending on the approach being discussed. Consequently, we will begin by considering exactly what a cognitive architecture does entail in the two different approaches: cognitivist and emergent. Following that, we will consider the necessary and desirable features of a cognitive architecture before embarking on a survey of specific cognitive architectures.

5.2.1 The Cognitivist Perspective on Cognitive Architectures

In the cognitivist paradigm, the focus in a cognitive architecture is on the aspects of cognition that are constant over time and that are independent of the task [128, 214, 217, 309]. In Sun's words [364]:

> "a cognitive architecture is a broadly-scoped domain-generic computational cognitive model, capturing the essential structure and process of the mind, to be used for broad, multiple-level, multiple-domain analysis of behaviour."

Since cognitive architectures represent the fixed part of cognition, they cannot accomplish anything in their own right and need to be provided with or acquire knowledge to perform any given task. A cognitivist cognitive architecture is a generic computational model that is neither domain-specific nor task-specific. It is the knowledge which populates the cognitive architecture that provides the requisite specificity. This combination of a given cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model*.

In most cognitivist systems the knowledge incorporated into the model is normally determined by the human designer, although there is an increasing use of machine learning to augment and adapt this knowledge [217, 364].

A cognitive architecture specifies the overall structure and organization of a cognitive system, including the essential modules, the essential relations between these modules, and the essential algorithmic and representational details within these modules [364]. The architecture specifies the formalisms for knowledge representations and the types of memories used to store them, the processes that act upon that knowledge, and the learning mechanisms that acquire it. Typically, it also provides a way of programming the system so that intelligent systems can be instantiated in some application domain [214].

A cognitive architecture plays an important role in computational modelling of cognition in that it makes (or should make) explicit the initial set of assumptions upon which that model is founded. Sun notes that these assumptions can be derived from several sources: biological or psychological data, philosophical arguments, or ad hoc working hypotheses inspired by, e.g., neurophysiology, psychology, or computational models [364]. A cognitive architecture also provides a comprehensive framework for developing these ideas further.

5.2.2 The Emergent Perspective on Cognitive Architectures

For emergent approaches to cognition, which focus on development from a primitive state to a fully cognitive state over the life-time of the system, the architecture of the system is equivalent to its phylogenetic configuration: the initial state from which it subsequently develops. With emergent approaches, the need to identify an architecture arises from the intrinsic complexity of a cognitive system and the need to provide some form of structure within which to embed the mechanisms for perception, action, adaptation, anticipation, and motivation that enable the ontogenetic development over the system's life-time. It is this complexity that distinguishes an emergent developmental cognitive architecture from, for example, a connectionist robot control system that typically learns associations for specific tasks [186]. Again, the cognitive architecture of an emergent system corresponds to the innate resources and capabilities that are endowed by the system's phylogeny and which don't have to be learned but of course which may be developed further. These resources facilitate the system's ontogensis. They represent the initial point of departure for the cognitive system and they provide the basis and mechanism for its subsequent autonomous development, a development that may impact directly on the architecture itself. As we have stated already, the autonomy involved in this development is important because it places strong constraints on the manner in which the system's knowledge is acquired and by which its semantics are grounded (typically by autonomy-preserving anticipatory and adaptive skill construction) and by which an inter-agent epistemology is achieved (the subjective outcome of a history of shared consensual experiences among phylogenetically-compatible agents); see Table 5.1.

The presence of innate capabilities in an emergent system does *not* imply that the architecture is necessarily functionally modular, i.e. that the cognitive system is comprised of distinct modules each one carrying out a specialized cognitive task. If a modularity is present, it may be because it develops this modularity through experience as part of its ontogenesis or epigenesis rather than being prefigured by the phylogeny of the system (e.g. see Karmiloff-Smith's theory of representational redescription, [189, 190]). Even more important, it does not necessarily imply that the innate capabilities are hard-wired cognitive skills as suggested by nativist psychology (e.g. see Fodor [98] and Pinker [292]).[7] At the same time, neither does it necessarily imply that the cognitive system is a blank slate, devoid of any innate cognitive structures as posited in Piaget's constructivist view of cognitive development [290];[8] at the very least there must exist a mechanism, structure, and organization which allows the cognitive system to be autonomous, to act effectively to some limited extent, and to develop that autonomy.

Finally, since the emergent paradigm sits in opposition to the two pillars of cognitivism — the dualism that posits the logical separation of mind and body, and the functionalism that holds that cognitive mechanisms are independent of the physical platform [105] — it is likely that the architecture will reflect in some way the morphology of the physical body of which it is embedded and of which it is an intrinsic part.

It is worth pausing here to note that the cognitivist and emergent perspectives differ somewhat on this issue of innate structure. While in an emergent system, the cognitive architecture *is* the innate structure, it is not necessarily so with a cognitivist system. Sun contends that "an innate structure can, but need not, be specified in an initial architecture" [363]. He argues that an innate structure does not have to be specified or involved in the computational modelling of cognition and that architectural detail may indeed result from ontogenetic development. However, he concedes that non-innate structures should be avoided as much as possible and that we should adopt a minimalist approach: an architecture should include only minimal structures and minimal learning mechanisms which should be capable of "bootstrapping all the way to a full-fledged cognitive model".

5.2.3 Desirable Characteristics of a Cognitive Architecture

In his *Desiderata for Cognitive Architectures* [363], Sun identifies several desirable features of a cognitive architecture. These are

1. Ecological realism;
2. Bio-evolutionary realism;

[7] More recently, Fodor [99] asserts that modularity applies only to local cognition (e.g. deciding to take a ride on your bicycle) but not global cognition (e.g. planning to train for the Race Across America).

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

3. Cognitive realism;
4. Inclusivity of prior perspectives.[9]

The key idea behind *ecological realism* is that a cognitive architecture should focus on allowing the cognitive system to operate in its natural environment, engaging in "everyday activities". This means it has to be able to deal with being embodied and the attendant natural constraints on its actions and perceptions. It also means that the architecture has to deal with many concurrent and often conflicting goals with many environmental contingencies. Since human intelligence evolved from the capabilities of earlier primates, *bio-evolutionary realism* asserts that a cognitive model of human intelligence should be reducible to a model of animal intelligence. *Cognitive realism* means that a cognitive architecture should capture the essential characteristics of human cognition as we understand them from the perspective of psychology, neuroscience, and philosophy. Finally, the design of a cognitive architecture should include *prior perspectives* and capabilities. In other words, new models should draw on, subsume, or supercede older models.

Sun also elaborates on the behavioural and cognitive characteristics which should ideally be captured by a cognitive architecture and exhibited by a cognitive system.

From a behavioural perspective, a cognitive system should act and react without employing excessively complicated conceptual representations and extensive computation devoted to working through alternative strategies. That is, the system should behave in a "direct and immediate" manner. Furthermore, a cognitive system should operate sequentially, one step at a time, in a temporally-extended sequence of actions. This leads naturally to the characteristic of gradually-learned routine behaviours, typically acquired through a process of trial-and-error adaptation.

From the perspective of cognitive characteristics, Sun suggests that a cognitive architecture should comprise two distinct types of process: one explicit, the other implicit. The explicit processes are accessible and precise whereas the implicit ones are inaccessible and imprecise. Furthermore, there should be a synergy borne of interaction between these two types of process. There are, for example, explicit and implicit learning processes and these interact. The most important type of learning in a cognitive architecture is what Sun refers to as bottom-up learning whereby implicit learning is followed by explicit learning. In Sun's own work [362, 364], implicit processes operate on connectionist representations and implicit processes on symbolic representations. Thus, he adopts a strong hybrid approach to cognitive architectures. Finally, Sun argues for a form of modularity in a cognitive architecture so that some cognitive faculties are specialized and separate, either as functionally-encapsulated modules or as physically — neurophysiologically — encapsulated modules.

Langley, Laird, and Rogers [217] catalogue nine functional capabilities that should be exhibited by an ideal cognitive architecture. Although they focus mainly on cognitivist cognitive architectures in their examples, the capabilities they discuss also apply for the most part to emergent systems. The nine capabilities are:

1. Recognition and Categorization;
2. Decision Making and Choice;

[9] Sun refers to this inclusivity as "eclecticism of methodologies and techniques" [363].

3. Perception and Situation Assessment;
4. Prediction and Monitoring;
5. Problem Solving and Planning;
6. Reasoning and Belief Maintenance;
7. Execution and Action;
8. Interaction and Communication;
9. Remembering, Reflection, and Learning.

Let's look at each of these in turn.

Recognition and Categorization: A cognitive architecture must be able to recognize objects, situations, and events as instances of known patterns and it must be able to assign them to broader concepts or categories. A cognitive architecture should be able to learn new patterns and categories, modify existing ones, either by direct instruction or by experience.

Decision Making and Choice: Since a cognitive architecture exists to support the actions of a cognitive agent, it must provide a way to identify and represent alternative choices and then decide which are the most appropriate and select an action for execution. Again, an ideal cognitive architecture should be able to improve its decisions through learning.

Perception and Situation Assessment: A cognitive architecture must have some perceputal capacity and, since a cognitive agent typically has limited computational resources, it must have an attentive capacity to decide how to allocate these resources and to detect what is immediately relevant.

Prediction and Monitoring: A cognitive architecture should have some mechanism to predict future situations and events, typically based on an internal model of the cognitive agent's environment. Ideally, a cognitive architecture should have a mechanism to learn these models from experience and improve them over time.

Problem Solving and Planning: To achieve goals, a cognitive architecture must have some capability to plan actions and solve problems. A plan requires some representation of a partially-ordered sequence of actions and their effects. Problem solving differs from planning in that it may also involve physical change in the agent's world. As always, an ideal cognitive architecture should be able to deploy learning to support both planning and problem solving.

Reasoning and Belief Maintenance: The knowledge that complements a cognitive architecture constitutes the agent's beliefs about itself and its world, while planning is focussed on using this knowledge to effect some action and achieve a desired goal. The cognitive architecture should also have a reasoning mechanism which allows the cognitive system to draw inferences from these beliefs, either to maintain the beliefs or to modify them.

Execution and Action: Langley, Laird, and Rogers highlight the fact that "cognition occurs to support and drive activity in the environment". Consequently, a cognitive architecture must have some mechanism to represent and store motor skills that can be used in the execution of an agent's actions. As before, an ideal cognitive architecture will have some way of learning these motor skills from instruction or experience.

Interaction and Communication: A cognitive architecture should be able to communicate with other agents so that they can obtain and share knowledge. This may also require a mechanism for transforming the knowledge from internal representations to a form suitable for communication.

Remembering, Reflection, and Learning: Langley, Laird, and Rogers remark that it may also be useful for a cognitive archtitecture to have additional capabilities which are not strictly necessary but which may improve the operation of the cognitive agent. These are referred to as meta-management functions [347] and they are concerned with remembering (storing and recalling) the agent's cognitive experiences and with reflecting on them, for example, to explain decisions, plans, actions in terms of the cognitive steps that led to them. They include also a form of learning that is capable of generalizing from specific experiences of the cognitive system.

These nine capabilities are advocated by others, e.g. Sun who identifies at least twelve similar requisite functionalities in a cognitive architecture: perception, categorization, multiple representations, multiple types of memory, decision making, reasoning, planning, problem solving, meta-cognition, communication, action control and execution, and several types of learning (which will often be embedded in the other functional capabilities) [364]. Sun also notes that very few cognitive architectures support these functionalities fully. He stresses the importance of the interconnectivity between these processes and the dynamic interaction that arises as the cognitive system experiences and acts in its environment. In Sun's words: "we need to strive for complex[10] cognitive architectures that capture dynamics of cognition through capturing its constituent elements."

Langley, Laird, and Rogers conclude their paper by highlighting a number of challenges in cognitive architecture research [217]. These include mechanisms for selective attention, processes for categorization, support for episodic memory and processes to reflect on it, support from multiple knowledge representation formalisms, the inclusion of emotion in cognitive architectures to modulate cognitive behaviour, and the impact of physical embodiment on the overall cognitive process, including the agent's internal drives and goals.

Sun too identifies several challenges in designing cognitive architectures [364]. He warns of the perils of designing excessively-complicated models. A cognitive architecture should involve only what is "minimally necessary". Like Einstein, who believed "a scientific theory should be as simple as possible, but no simpler", Sun advocates that a cognitive architecture should be well constrained with as few parameters as possible, without compromising its broad-based domain-generic objectives [364]. He also notes that the validation of cognitive architectures poses a major but essential challenge and he highlights the need to guard against over-stating the case for any particular architecture: "as in any other scientific fields, painstakingly detailed work needs to be carried out before sweeping claims can be made" [364].

In designing a cognitive architecture, Sloman and his co-workers advocate a three-step process [138]. First, the requirements of the architecture should be

[10] Complex, but not excessively complex: see the second-next paragraph.

identified, partly through an analysis of several typical scenarios in which the eventual agent would demonstrate its competence. These requirements are then used to create an *architecture schema*: "a task and implementation independent set of rules for structuring processing components and information, and controlling information flow". This schema leaves out much of the detail of the final design choices, detail which is finally filled in by an instatiation of the architecture schema in a cognitive architecture proper on the basis of a specific scenario and its attendant requirements.

While not specifically targetting cognitive architectures, Krichmar et al. identify six design principles for systems that are capable of development [203, 206, 204]. Although they present these principles in the context of their brain-based devices, most are directly applicable to emergent systems in general. First, they suggest that the architecture should address the dynamics of the neural element in different regions of the brain, the structure of these regions, and especially the connectivity and interaction between these regions. Second, they note that the system should be able to effect perceptual categorization: i.e. to organize unlabelled sensory signals of all modalities into categories without a priori knowledge or external instruction. In effect, this means that the system should be autonomous and, as noted by Weng [394], p. 206, a developmental system should be a model generator, rather than a model fitter (e.g. see [278]). Third, a developmental system should have a physical instatiation, i.e. it should be embodied, so that it is tightly coupled with its own morphology and so that it can explore its environment. Fourth, the system should engage in some behavioural task and, consequently, it should have some minimal set of innate behaviours or reflexes in order to explore and survive in its initial environmental niche. From this minimum set, the system can learn and adapt so that it improves[11] its behaviour over time. Fifth, developmental systems should have a means to adapt. This implies the presence of a value system (i.e. a set of motivations that guide or govern it development). These should be non-specific (in the sense that they don't specify what actions to take) modulatory signals that bias the dynamics of the system so that the global needs of the system are satisfied: in effect, so that its autonomy is preserved or enhanced. Such value systems might possibly be modelled on the value system of the brain: dopaminergic, cholinergic, and noradrenergic systems signalling, on the basis of sensory stimuli, reward prediction, uncertainty, and novelty. Krichmar et al. also note that brain-based devices should lend themselves to comparison with biological systems.

5.2.4 A Survey of Cognitive Architectures

For the remainder of the chapter, the term cognitive architecture will the taken in the general and paradigm non-specific sense. By this we mean the minimal configuration of a system that is necessary for the system to exhibit and develop cognitive capabilities and behaviours: the specification of the components in a cognitive system, their function, and their organization as a whole.

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

Appendix A contains a synopsis of twenty cognitive architectures spanning the cognitivist, emergent, and hybrid paradigms. The cognitive architectures surveyed are Soar, EPIC, ACT-R, ICARUS, ADAPT, GLAIR, and CoSy Architecture Schema (cogntivist); Autonomous Agent Robotics, Global Workspace, I-C SDAL, SASE, Darwin, Cognitive-Affective (emergent); and HUMANOID, Cerebus, Cog: Theory of Mind, Kismet, LIDA, CLARION, PACO-PLUS (hybrid). 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.

Table 5.2 shows a comparison of the twenty architectures surveyed vis-à-vis a subset of the fourteen characteristics of cognitive systems which we discussed in Sect. 5.1. We have omitted the first five and last two characteristics — Computation Operation, Representational Framework, Semantic Grounding, Temporal Constraints, and Inter-agent Epistemology, Cognition, and Philosophical Foundation — because these can be inferred directly by the paradigm in which the system is based: cognitivist, emergent, or hybrid. The seven remaining characteristics — embodiment, perception, action, anticipation, adaptation, motivation, and autonomy — are crucial for enactive cognitive systems which we are adopting as our framework for development, as discussed in Chap. 1.

Table 5.2 reveals a number of interesting points.

The cognitivist cognitive architectures typically address only a subset of the seven characteristics. Only one — GLAIR [339] (see Sect. A.1.6) — addresses autonomy, only one — CoSy Architecture Schema [137, 138] (see Sect. A.1.7) — addresses motivation, and only one — ADAPT [30] (see Sect. A.1.5) — makes any strong commitment to embodiment. To an extent, this is not surprising, given the functionalist and dualist foundation of cognitivism: a mind that is divorced in principle from its body doesn't need to worry about embodiment, survival and the preservation of autonomy, or the motivations that modulate that autonomy-preserving process. Most of the cognitivist architectures do address perception, action, anticipation, and adaptation, although it is significant that only ICARUS [60, 213, 214, 215, 216] (see Sect. A.1.4) addresses adaptation in the strong sense that entails development through the creation of new representational frameworks or models. In the specific case of ICARUS, the cognitive architecture can learn hierarchically-structured skills to solve new problems.

All of the emergent cognitive architectures address most of the seven characteristics. Three of the seven architectures surveyed address all seven. Again, this is not surprising, given the foundations of the emergent paradigm in self-organizing autonomy-preserving embodied systems. All address embodiment, perception, action, and autonomy to a lesser or greater extent. They differ in whether or not they target anticipation and adaptation. Only the Global Workspace cognitive architecture [335, 336, 337, 338] (see Sect. A.2.2) and the Cognitive-Affective schema [264, 415] (see Sect. A.2.6) address anticipation in depth and only the SASE architecture [394, 395, 393] (see Sect. A.2.4) and the Cognitive-Affective schema [264, 415] (see Sect. A.2.6) address adaptation in a strong manner. The Global Workspace architecture uses internal simulation of interaction with the environment

Table 5.2 Cognitive architectures vis-à-vis the seven of the fourteen characteristics of cognitive systems. Key: '×' indicates that the characteristic is strongly addressed in the architecture, '+' indicates that it is weakly addressed, and a space indicates that it is not addressed at all in any substantial manner. A '×' is assigned under the heading of Adaptation only if the system is capable of development (in the sense of creating new representational frameworks or models) rather than simple learning (in the sense of model parameter estimation). 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.

Cognitive Architecture	Embodiment	Perception	Action	Anticipation	Adaptation	Motivation	Autonomy
Cognitivist							
Soar				+	+		
Epic		+	+	+			
ACT-R		+	+	+	+		
ICARUS		+	+	+	×		
ADAPT	×	×	×	+	+		
GLAIR		+	+		+		+
CoSy		+	+		+	+	
Emergent							
AAR	×	×	×			+	×
Global Workspace	+	+	+	×		×	×
I-C SDAL	+	+	+	+	+	×	×
SASE	×	×	×	+	×	×	×
Darwin	×	×	+		+	×	×
Cognitive-Affective	×	×	×	×	×	×	×
Hybrid							
HUMANOID	×	×	×	+	+	+	
Cerebus	×	×	×	+	+		
Cog: Theory of Mind	×	×	×	+			
Kismet	×	×	×			×	
LIDA	+	+	+	×	×	+	+
CLARION		+	+	×	×	+	+
PACO-PLUS	×	×	×		×		

to effect anticipation and planning, with action selection being modulated by affective motivation mechanisms. Significantly, it is the concurrent operation of the components / sub-systems of the cognitive architecture as they compete for access to a global workspace that governs the behaviour of the system, with the resultant behaviour emerging as a sequence of states arising from their interaction. It is noteworthy that, reflecting contemporary thinking in neuroscience, in the Cerebus cognitive architecture [170, 171] (see Sect. A.3.2) each of these components / sub-systems maintains its own separate and limited representation of the world and the task at hand: different motor effectors need different sensory inputs, derived in

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

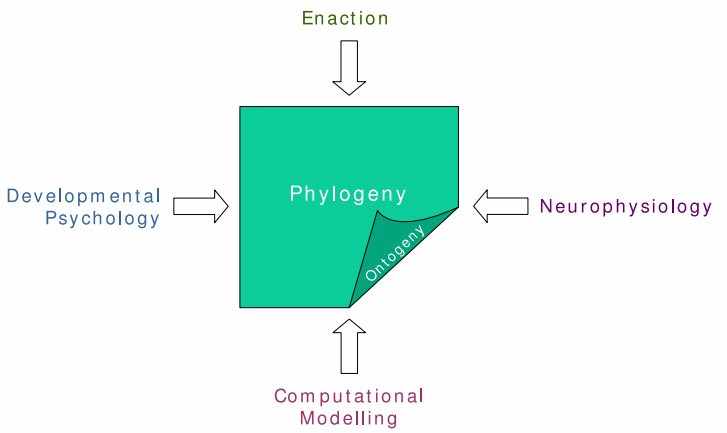


Fig. 6.1 Four influences contributing to the phylogentic and ontogenetic principles of developmental cognitive systems

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.

6.1 Phylogeny

6.1.1 Guidelines from Enaction

Chapter 1 yields the following nine guidelines.[1]

1. The system should incorporate a rich array of physical sensory and motor interfaces which allow the system to act on the world and perceive the effects of these actions.
2. The 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.
3. The system requires a humanoid morphology if it is to construct an understanding of its environment that is compatible with that of human cognitive agents.
4. 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.
5. 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.
6. 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.
7. The system should incorporate processes for the generation of knowledge effected by learning affordances whereby the perception of an object is interpreted as affording the opportunity for the system to act on it in a specific way with a specific outcome.
8. 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.
9. The system should also incorporate processes for grounding internal simulations in actions to establish by observation their validity.

6.1.2 Guidelines from Developmental Psychology

Chapter 2 yields the following fourteen guidelines. The last nine of these are concerned with pre-configured core abilities to enable knowledge generation relating to the perception of (a) objects and their movements, (b) spatial relationships between objects, (c) numbers of objects, (d) persons and their actions. These core knowledge

[1] Chapter 1 and these nine guidelines are based directly on a study of enaction as a framework for development in cognitive robotics [385].

systems should persist as domain-specific, task-specific, and encapsulated capabilities so that each sub-system performs a cohesive function relatively independently of other systems. However, they should also act as building blocks for scaffolding new cognitive abilities and more complex cognitive tasks by recruiting existing core knowledge systems in new ways.

10. Movements should be organized as actions. Actions are planned: directed by goals, guided by prospection, and triggered by motives.
11. The system should have at least two primary motives that drive actions, one social and one explorative:
    - The social motive should manifest as a fixation on social stimuli, imitation of basic gestures, and engagement in social interaction (e.g. turn-taking).
    - The explorative motive is concerned with finding out about the system's own action capabilities and the expansion of its space of actions.
12. Attention should be fixated on the goal of an action, not on its constituent movements.
13. Morphology should be integral to the model of cognition and changes in morphology should involve matching changes in the perceptual system to improve the extraction of information for controlling specific actions.
14. Pre-structured sensory-motor couplings exist at birth. Early in ontogenesis, movements should be constrained to reduce the number of degrees of freedom and thereby simplify the control task.
15. The system should divide up its optic array into regions that exhibit five characteristics:
    - inner unity;
    - a persistent outer boundary;
    - cohesive and distinct motion;
    - relatively constant size and shape when in motion;
    - a change in the behaviour or motion of one or both of the objects when contact occurs with another object.

    These regions are perceived as objects
16. 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. 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.
17. Navigation should be based on representations that are dynamic and ego-centric rather than eco-centric. Navigation should use path integration, navigating by moving from place to place, re-orienting as you go.
18. Re-orientation should be 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.

19. The system should be attracted to people and especially to their faces, their sounds, movements, and features.
20. The system should pay preferential attention to biological motion rather than non-biological mechanical motion.
21. The system should recognize people and expressions and perceive the goal-directioned nature of actions.
22. The system should gaze longer when the person looks directly at it.
23. The system should perceive and communicate emotions by facial gesture and engage in turn-taking.

6.1.3 Guidelines from Neurophysiology

Chapter 4 yields the following six guidelines from neurophysiology. It is likely to be advantageous to design an artificial cognitive system along similar lines.

24. The system should encode space in several different ways, each of which is specifically concerned with a particular motor goal.
25. The motor system should encode a repertoire of goal-oriented actions (rather than just the component movements that constitute them) with the goal of the action being provided by associated effector-specific percepts. This is a pre-requisite for learning affordances.
26. 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.
27. 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).
28. Pre-motor theory of attention — spatial attention: 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.
29. Premotor theory of attention — selective attention: the preparation of a motor program in readiness to act on specific objects should predispose the perceptual system focus attention on those objects.

6.1.4 Guidelines from Computational Modelling

As we saw in Chap. 5, computational modelling provides several guidelines for the design of a cognitive architecture which adheres to the enactive systems approach. Since many aspects of the cognitive architectures that were surveyed when drawing up these guidelines are based on the principles of enactive systems as well as developmental psychology and neurophysiology, there are inevitably several

guidelines that are similar to those we have already discussed. Nonetheless, we include them here for completeness as they often serve to crystalize the point at issue. For reference, we note when a guideline is similar to one already listed. A cognitive architecture acting as the phylogenetic basis for development should have the following fourteen characteristics.

30. The system should have a minimal set of innate behaviours for exploration and survival, i.e. preservation of autonomy (cf. Guideline No. 2).
31. The system should have a value system — a set of task non-specific motivations — that guide or govern actions and development (cf. Guideline Nos. 6 and 11).
32. The system should have an attentional mechanism (cf. Guideline Nos. 12, 20, 22, 28, and 29).
33. The system should learn from experience the motor skills associated with actions (cf. Guideline No. 14).
34. 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 (cf. Guideline No. 2).
35. The system should anticipate and plan based on internal simulation of interaction with the environment. (cf. Guideline Nos. 8 and 26).
36. Action selection should be modulated by affective motivation mechanisms (cf. Guideline Nos. 2 and 10).
37. The system should have separate and limited representations of the world and the task at hand in each component / sub-system (cf. Guideline No. 24).
38. The system should have both transient and generalized episodic memories of past experiences (cf. Guideline No. 18).
39. The system should have a procedural memory of actions and outcomes associated with episodic memories (cf. Guideline Nos. 7 and 17).
40. The system should have the ability to learn based on comparison of expected and observed outcomes of selected actions, resulting in either generalization or refinement of the associated action (cf. Guideline No. 9).
41. The system should have the ability to learn and use co-joint object-action affordances by exploration (cf. Guideline Nos. 7 and 25).
42. The system should have hierarchically-structured representations for the acquisition, decomposition, and execution of action-sequence skills (cf. Guideline No. 27).
43. 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 (cf. Guideline No. 24).

These forty-three guidelines are summarized in Table 6.1 under the four headings denoting their origin: Enaction, Developmental Psychology, Neurophysiology, and Computational Modelling.

Table 6.1 A summary of the guidelines for the configuration of the phylogeny of a humanoid robot that is capable of developing cognitive abilities

Guidelines for the Phylogeny of a Developmental Cognitive System	
Number	**Guideline**
	Enaction
1	Rich array of physical sensory and motor interfaces
2	Autonomy-preserving processes of homeostasis
3	Humanoid morphology
4	Self-modifi cation to expand actions and improve prediction
5	Autonomous generative model construction
6	Internal social and exploratory motives
7	Learning affordances
8	Internal simulation to predict, explain, & imagine events, and scaffold knowledge
9	Grounding internal simulations in actions
	Developmental Psychology
10	Movements organized as actions
11	Social and explorative motives
12	Attention fi xated on the goal of an action
13	Morphology integral to the model of cognition
14	Early movements constrained to reduce the number of degrees of freedom
15	Perception of objecthood
16	Discrimination & addition of small numbers; groups of large numbers
17	Navigation based on dynamic ego-centric path integration
18	Re-orientation based on local landmarks
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
	Neurophysiology
24	Encode space in motor & goal specifi c manner
25	Motor system encoding of actions with associated effector-specifi c percepts
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
	Computational Modelling
30	Minimal set of innate behaviours for exploration and survival
31	Value system that govern actions and development
32	Attentional mechanism
33	Learn from experience the motor skills associated with actions
34	Affective drives associated with autonomy-preserving processes of homeostasis
35	Anticipation and planning based on internal simulation
36	Action selection modulated by affective motivation mechanisms
37	Separate representations associated with each component / sub-system
38	Transient and generalized episodic memories of past experiences
39	Procedural memory of actions and outcomes associated with episodic memories
40	Mechanism to learn based on comparison of expected and observed outcomes
41	Mechanism to learn co-joint object-action affordances by exploration
42	Hierarchically-structured representations of action-sequence skills
43	Concurrent competitive operation of components and subsystems

6.1.5 A Summary of the Phylogenetic Guidelines for the Development of Cognition in Artificial Systems

And so, based on the foregoing phylogenetic guidelines, what conclusions can we draw? To answer this question, we first re-group the forty-three guidelines under the seven headings we used in Chap. 5 to characterize cognitive architectures:

1. Embodiment
2. Perception
3. Action
4. Anticipation
5. Adaptation
6. Motivation
7. Autonomy

These re-grouped guidelines are summarized in Table 6.2.

We can now draw seven broad conclusions.

First, developmental cognitive systems have to be embodied in humanoid form if the epistemological understanding of the developed systems is required to be consistent with that of humans. What is also clear is that the complexity and sophistication of the cognitive behaviour is dependent on the richness and diversity of the coupling and therefore the potential richness of the system's actions.

Second, perceptual mechanisms should isolate regions of sensory stimuli displaying the characteristics of objecthood, discriminating between small numbers of objects, and distinguishing objects using motor information. There should be mechanisms that allow the construction of hierarchically-organized percepts. The system should have an attentional system which fixates on the goals of actions, which is attracted to people and faces, which responds strongly to mutual gaze, and which reflects the pre-motor theory whereby attention is modulated by prepared motor programs acting either spatially ("attention to where") or selectively ("attention to what").

Third, the system's actions are guided by prospection, directed by goals, and triggered by affective motives. They are initially constrained in their numbers of freedom and the motor-programs that constitute them are learned. There should be mechanisms that allow the construction of hierarchical action sequences. Navigation should be effected dynamically using ego-centric representations of geometric landmarks.

Fourth, because cognitive systems are anticipatory and prospective, it is crucial that they have or develop some mechanism to rehearse hypothetical scenarios through some process of internal simulation in order to predict, explain, and imagine events. There should be a mechanism to use this outcome to modulate the behaviour of the system. Internal simulation is also used to scaffold new knowledge through the developmental generative model-building processes. These processes should incorporate transient and generalized episodic memories of events and a procedural memory that links actions to perceptions. Again, there should be mechanisms that facilitate hierarchical representations (episodic, procedural, or both).

Table 6.2 A summary of the guidelines for the configuration of the phylogeny of a humanoid robot regrouped under the seven headings used in Chap. 5 to characterize cognitive architectures. 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**
	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 fixated 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-modification 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 specific 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

Fifth, a developmental cognitive architecture must be capable of adaptation and self-modification, both in the sense of parameter adjustment of phylogenetic skills through learning and, more importantly, through the modification of the very structure and organization of the system itself so that it is capable of altering its system dynamics based on experience, to expand its repertoire of actions, and thereby adapt to new circumstances and the enhance its prospective capabilities. The focus of development should be on generative model construction, bootstrapped by learned affordances.

Sixth, this development should be driven by both explorative and social motives, the first concerned with both the discovery of novel regularities in the world and the potential of the system's own actions, the second with inter-agent interaction, shared activities, and mutually-constructed patterns of shared behaviour.

Finally, a developmental cognitive system should be constituted by a network of competing and cooperating distributed multi-functional sub-systems (or cortical circuits), each with its own limited encoding or representational framework, together achieving the cognitive goal of effective behaviour realized through autonomy-preserving homeostasis. This network forms the system's phylogenetic configuration and its innate abilities.

6.2 Ontogeny

6.2.1 Guidelines from Developmental Psychology

Chapter 3 provides us with an overview of the key cognitive abilities that are developed early on in infants. In this section, we wish to complement the insights on the phylogeny of the system from Chaps. 1, 2, and 4 by providing a similar list of insights into the ontogeny of the system from Chap. 3 and, specifically, to identify those capabilities that result from this development, focussing in particular on early development. These are listed in Table 6.3. We will then propose a procedure for this development through a series of scenarios that are realized by scripted experiments. We won't concern ourselves here with the exact timeline of development since this has already been discussed in depth in Chap. 3.

Before we embark on this exercise, let's remind ourselves again of some of the key messages from Chap. 3 on the foundations of 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: (a) the discovery of novelty and regularities in the world, and

(b) the discovery of the potential of the infant's own actions. In the development of perception, there are two processes: (a) the detection of structure or regularity in the flow of sensory data, and (b) the selection of information which is relevant for guiding action.

6.2.2 Scenarios for Development

Table 6.3 identies four primary areas of early development based on the psychology of human infants: vision, posture, gaze, and reaching & grasping. Using these, we now present a set of scenarios for development which are then fleshed out in the next section as a set of scripted practical exercises. First, we make some general remarks.

The primary focus of the early stages of ontogenesis is to develop manipulative action based on visuo-motor mapping, learning to decouple motor synergies (e.g. grasping and reaching), anticipation of goal states, learning affordances, interaction with other agents through social motives, and imitative learning. Needless to say, ontogenesis and development are progressive. In the following, we emphasize the early phases of development, building on the phylogenetic skills outlined in the Sect. 6.1 and scaffolding the cognitive abilities of the robot to achieve greater prospection and increased (action-dependent) understanding by the robot of its environment and other cognitive agents.

It is important to emphasize that the development program that we propose to facilitate the ontongenesis of the robot is biologically inspired and tries to be as faithful as possible to the ontogenesis of neonates. Consequently, the development of manipulative action will build primarily on visual-motor mapping.

The following are the scenarios that will be used to provide opportunities for the robot to develop, in order of their deployment over time.

Scenario 1: Reaching for Objects

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

Scenario 2: Grasping Objects

Once the robot masters the extension of the hand towards objects in the surrounding and can flex the fingers around them, grasping skills can develop. However, the robot must have some kind of motive for grasping objects in order to make this happen. Note that it is the sight of the object that should elicit anticipations

of the sensory consequences of the action. Infants who are at the transition to mastering the grasping of objects anticipate crudely the required orientation of the hand. They open the hand fully when approaching any object which optimizes the chances of getting the object into the hand. Adjusting the opening of the hand during the approach to the size of the object to be grasped develops as the infant becomes experienced with object manipulation. The timing of the

Table 6.3 A summary of the capabilities that are the subject of early development in a human infant. These form the basis for the ontogenesis of a humanoid robot.

Principal Capabilities Subject to Early Development
Vision
Visual acuity
Colour processing
Ocular convergence for objects beyond 20 cm
Depth perception from binocular stereo disparity
Object discrimination based on motion information
Smooth pursuit
Depth perception from motion parallax
Ability to perceive binocular depth
Ocular convergence for reaching
Depth perception of looming objects
Depth perception based on perspective, size, interposition, shading
Object discrimination based on colour
Posture
Head stabilization when lying prone and lifting the head
Head stabilization when sitting
Anticipatory adjustment of posture when reaching
Gaze
Vestibular gaze stabilization to compensate for head movement
Saccadic eye movements, ability to engage and disengage attention
Limited smooth pursuit ability
Attentional processes are present: gaze directed toward attractive objects and novel appearances or events
Infants achieve adult level of smooth pursuit
Reaching and Grasping
Visual control of arm but no control of fingers for grasping
Arm and finger motions governed by global extension/flexion synergies
Hand is fisted when extending the arm
Open hand when reaching, but only when visually guided
Hand closing when close to object
Reaching and grasping as a function of object properties
Adjustment of hand size when reaching
Hand closes when in the vicinity of object
Differentiated finger grasping, e.g. pincer grasp
Grasping starts when reaching: i.e. one integrated reach-grasp act

grasp is controlled visually but, to begin with, at the expense of interrupting the flow of the action (the movement is temporarily stopped before the close around it). This coordination also improves as a function of experience.

Scenario 3: Affordance-based Grasping

Grasping objects as a function of their use only develops after infants master reaching and grasping objects in a versatile way towards the end of the first year of life. The first manipulative actions are general and explorative: squeezing, turning, shaking, putting into the other hand etc. The purpose can be said to learn about object properties. More specific and advanced object manipulation skills only develop after the end of the first year of life, such as putting objects into apertures, inserting one object into another, position lids on pans, building towers of blocks. Mastering actions like these relies on anticipation of goal states of manipulatory actions. This is how we intend the robot to develop its manipulatory action. The sensory effects of manipulatory action should be primarily visual, like the disappearance of the object into the hole.

Scenario 4: Imitative Learning

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

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

The robot starts with an innate visual-motor map that enables it to get its hand into the visual field. Thus, the robot also needs to have an innate conception of space in motor coordinates. When the hand is in the visual field, the robot tries to maintain it there. The robot should also be able to move its hand towards graspable objects in the visual field. In order to do all this, the robot should be equipped with motives to move the hand into the visual field and towards objects that can be grasped.

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

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

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

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

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

6.2.3 Scripted Exercises

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

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

6.2.3.1 Looking

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

Saccades and gaze redirection

A face pattern is introduced into the peripheral visual field (30° from the centre). The visual angle corresponds to that of a real face at 0.5m. When this happens, the robot moves the eyes and head to position the face at the centre of the visual field. They both start at the same time, but the eyes arrive first to its new position. When the eyes are at the final position and the head moves there, the gaze stays at the fixation object while the eyes counter rotate until they look straight ahead again. The same thing should also happen when a colourful object (3° – 8° visual angle) is introduced into the visual field or when a sounding object is introduced

to the side of the robot (30° – 50°). New objects that the robot has not seen before will attract the gaze more than familiar objects.

Gaze redirection and fixation

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

Saccades, gaze redirection, and dynamic fixation (tracking)

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

Minimization of saccade correction by learning: tracking through occlusion

An object moves in the visual field and gets temporarily occluded behind some other objects. The robot stops the eyes at the disappearance point and then makes a saccade to the other side of the occluder. The saccade will predict when and where the object will appear.

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

6.2.3.2 Reaching

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

Reaching towards a visual target (hand)

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

Reaching towards a visual target (body)

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

Reaching towards a visual target (moving object)

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

Learning efficient reaching & learning when not to reach

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

Again, some notes are in order. Turning the head toward the arm-hand as it enters the field of view is based on both visual and proprioceptive data. It implies a capability for hand detection and hand localization. The bimanual behaviour should be emergent. Moving the arm to different parts of the robot body and touching them implies both haptic and force feedback.

6.2.3.3 Reaching and Grasping

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

Reaching for a fixated static object

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

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

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

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

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

Bimanual manipulation and experimentation
After the object is grasped, the robot will examine the object by turning it around. Both hands will participate in this activity. One hand will hold the object in a fixed position while the other hand is moved over it in order to feel its surface and examine its interior. Through this activity the robot will built an object representation of familiar objects.

Hand-to-hand transfer
The object will be transferred from one hand to the other while the robot fixates the object (maybe also transferring repeatedly between the hands). The transfer should be as smooth and continuous as possible. This means that the delivering hand should let go of the object at the same time as the receiving hand grasps it.

Hand and arm object relocation to a fixation point via intermediate landmarks
After grasping an object, the robot will move it to another position and deposit it there. The robot will turn its gaze towards the goal position of the action while the

object is moved there. If the object is moved to its final position via an obstacle, the robot will fixate the obstacle and when the hand with the object has cleared the obstacle, the gaze will go to the final position.

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

6.2.3.4 Reach and Posture

Once these capabilities have been demonstrated, we move on to consider reaching and posture. In this case, the robot sits without support, exhibiting compensation for inertia and gravity, leaning forward, and using the other hand to counterbalance.

Similarly, the next stage in the development of the robot deals with postural control in action. Here, the robot sits independently and moves by crawling. The robot crawls and prepares a reach during crawling, it manages a transition from crawling to sitting, it balances while sitting, and it balances during action. The robot adjusts its posture to compensate for the action and the body is stabilized so when the robot grasps, the other hand counter-balances.

6.2.3.5 Object Containment

The next stage is to consider object containment. The robot sits independently in front of two objects, one of them is smaller than the other which is larger and hollow. The smaller object can be fitted into the larger object. The robot picks up one of the objects and inspects it visually from several viewpoints. The robot picks up the other object with the other hand and inspects it from several viewpoints. It then turns one of the object such that it fits into the other one.

6.2.3.6 Pointing and Gesturing

Finally, we consider pointing and gesturing. The robot sits in front of a human partner. An object is situated between them. The robot turns head and eyes toward the partners face and then towards the object and then towards the partner again. The robot then opens the hand with the palm up and moves the upper body forward as if wanting the partner to give it the object.

6.2.3.7 A Comprehensive Exercise

The following exercise is designed to demonstrate the integration of all the elements of the robot phylogeny.

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

1. The robot turns to look at one of the objects with head-eyes. It raises its right arm-hand and points to the attended object. It then assumes a crawling posture and crawls up to the objects. During the last stride the right arm is lifted (predictively).
2. When it arrives at the object, it assumes a sitting position, grasps the object and hands it to the human partner. This is repeated with the other object.
3. The human partner then picks up one of the objects and stretches it towards the robot who opens the hand and grasps the object.
4. After this, the human partner picks up the other object and hands it to the robot who transfers the first object to the other hand before receiving it.
5. Then the human partner turns his/her head and eyes toward one of the objects and points at it. The robot turns its head and eyes toward the same object. The human partner then extends one of its arms, points to the object and places the hand in a begging posture. The robot picks up the object and hands it to the human partner.
6. Now the human partner and the robot have one object each. The human partner picks up his/her object and drops it into one of two buckets. After this the robot picks up the other object and drops it into the other bucket (the gaze should move to the goal, not track the action).

This effectively completes the roadmap: a set of forty-three guidelines, four scenarios, and seven scripted experiements, all drawn from the principles of enactive cognitive systems, developmental psychology, neurophysiology, and computational modelling, and together determining the phylogeny and ontogeny of a humanoid robot that is capable of developing cognitive capabilities. In the next chapter, we discuss the progress that has been made to date in following this roadmap and, in particular, we address the design and implementation of a cognitive architecture for the iCub humanoid robot.

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

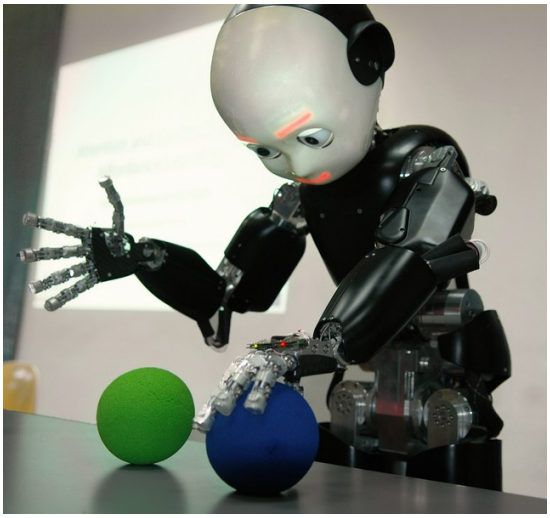


Fig. 7.1 The iCub humanoid robot: an open-systems platform for research in cognitive development

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.

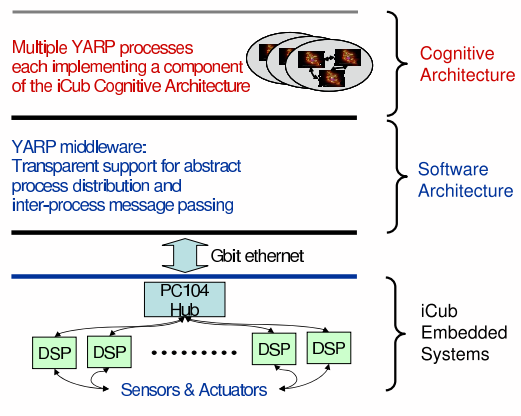


Fig. 7.3 The layers of the iCub architecture: the cognitive architecture and software architecture run on multiple computers located remotely from the iCub robot. The PC104 hub and the DSPs are located on the iCub. Communication between these two sub-systems is effected by a 1Gbit ethernet connection.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

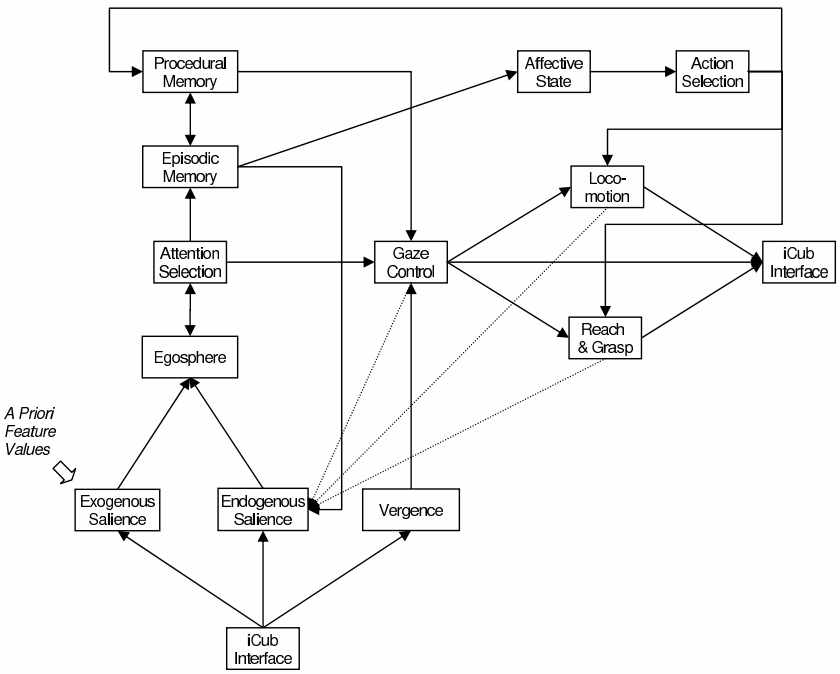


Fig. 7.4 The iCub cognitive architecture

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, eye-brow 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 ego-sphere, 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 ($\vartheta = 0^\circ$, $\varphi = 0^\circ$) 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*

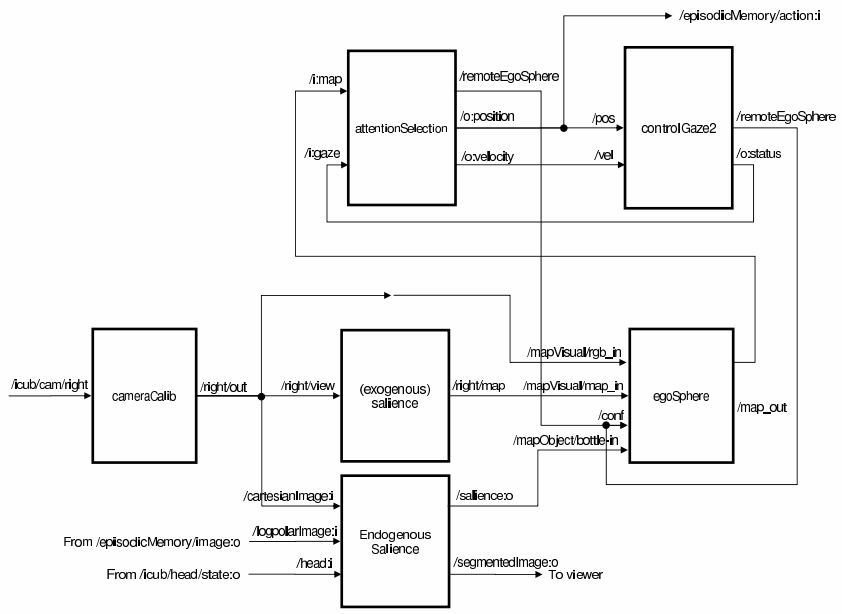


Fig. 7.5 Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part A)

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.

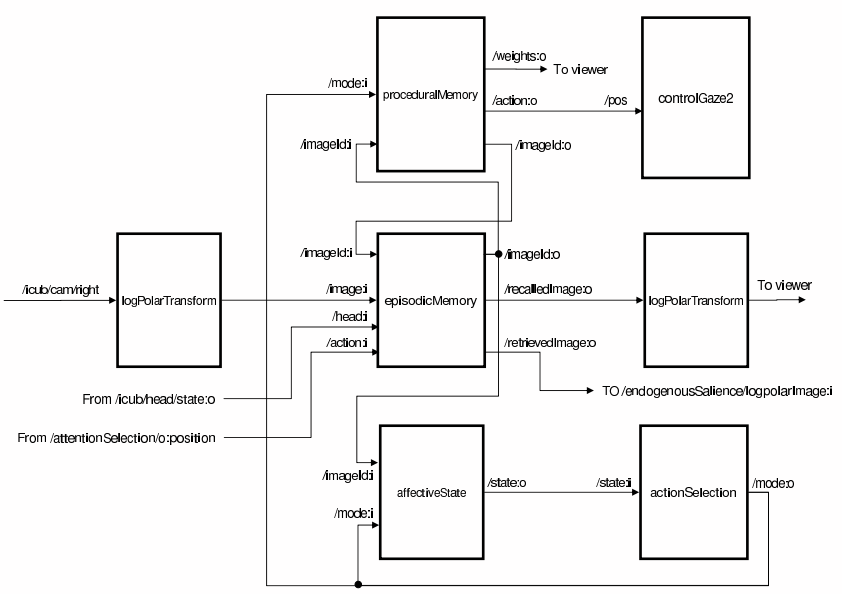


Fig. 7.6 Implementation of the iCub cognitive architecture, Version 0.4, as a YARP application (Part B)

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. $P_a, A_1, P_b, A_2, P_c, ..., P_z$.

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 $P, A, P, A, ...$ 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 fixated 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-modification 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 specific 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.

8.4 Multi-modal and Hierarchical Episodic Memory

As currently specified, the Episodic Memory component in the iCub cognitive architecture is a very simple representation of the iCub's experiences and there are several natural ways in which it could be extended or augmented. One obvious requirement, especially in the context of the cognitive architecture attention subsystem, is the need to include aural, tactile, and haptic information. One way to do this would be to extend the Episodic Memory to be multi-modal, with a composite multi-sensory storage and recall. This has the disadvantage of necessarily associating every modality with each data set, even though no significant sensory stimulus may be present for that percept (and vice versa). An alternative would be to effect an explicit episodic memory for each modality and link them through a hetero-associative memory. In addition, the robot's affective state might also be incorporated into this multi-modal framework so that emotions are associated with events. Ideally, such a schema would also encompass hierarchical representations for perceptual understanding and action execution. Again, this remains an imposing challenge for the future.

8.5 Generalization and Model Generation

Autonomous model generation is one of the hallmarks of enactive cognition. In the current implementation of the iCub cognitive architecture, all associations are based purely on the interaction history of the iCub and there is no capability for generalizing from these experiences or using them to imagine new ones. The iCub Episodic Memory and Procedural Memory components need to be augmented by some form of generalization. At present, the Episodic Memory simply does one-shot learning, while what is also needed is a memory — often referred to as semantic memory — that consolidates multiple encounters of the same experience. Together they form an explicit declarative memory, in contrast to implicit procedural memory which encapsulates temporal sequencing and skill-based learning. The question is whether this ability to generalize should be encapsulated or subsumed into the existing Episodic Memory. Neuroscientific evidence suggests it should not. For example, McClelland et al. have proposed that the hippocampal formation and the neocortex form a complementary system for learning [242]. The hippocampus facilitates rapid autoassociative and heteroassociative learning which is used to reinstate and consolidate learned memories in the neocortex in a gradual manner. In this way, the hippocampal memory can be viewed not just as a memory store but as a "teacher of the neocortical processing system". This suggests that the best way to proceed would be to implement a separate long-term semantic/generalized memory which takes as input the output of the current episodic memory. The experience gathered by reinforcement learning, encapsulated in the episodic memory, and periodically updating the long-term generalized memory thereby gives rise to an increased space of potential action.

What we have described is effectively a process of generative model building: extrapolating from experience and then validating, refining, or abandoning of the

extrapolated model on the basis of subsequent experience. The key is to accomplish this enactive sense-making by autonomous homeostatic processes, the goal of which is, as we have emphasized throughout this book, to extend the cognitive system's time horizon of anticipation and to increase its space of viable actions.

8.6 Homeostasis and Development

Development implies the progressive acquisition of anticipatory capabilities by a system over its lifetime through experiential learning. Development depends crucially on the motives which underpin the goals of actions. The two most important motives that drive actions and development are social and exploratory. There are at least two exploratory motives, one focussing on the discovery of novelty and regularities in the world, and one focussing on the potential of one's own actions. A challenge that faces all developmental embodied robotic cognitive systems is to model these motivations and their interplay, to identify how they influence action, and thereby build on the system's phylogeny through ontogenesis to develop every-richer cognitive capabilities.

For enactive systems, this challenge can be addressed by tackling the twin problems of homeostasis and self-modification (cf. Guidelines 2 & 4).[1] Since this homeostasis — autonomy-preserving self-regulation — entails structural determination, the homeostatic processes need to regulate the system's actions to ensure that the conditions required for the maintenance of autonomy are preserved in the environment. This, in turn, depends both on the system's internal structures and its physical realization, and both must figure in whatever homeostatic processes are embedded in the system's cognitive architecture. It also means that the conditions required for the maintenance of autonomy must be explicitly identified.

Since homeostasis is concerned with the maintenance of autonomy by structural coupling through the system's phylogenetic repertoire of actions and anticipatory capability, the space of environmental perturbations it can withstand is consequently limited. The purpose of self-modification is to develop the system so that it has a larger repertoire of actions and a greater degree of anticipation to enable it to withstand a larger space of perturbations by the environment. As Bickhard puts it when discussing recursive self-maintenant systems — systems that contribute actively to the conditions for persistence — these systems can deploy different processes of self-maintenance depending on environmental conditions: "they shift their self-maintenant processes so as to maintain self-maintenance as the environment shifts" [40]. Viewed in this way, development and self-modification are intrinsically linked to the processes of homeostasis, giving them more degrees of freedom in the manner in which autonomy is perserved. Perhaps the greatest challenge of all is to model the mechanisms and dynamics of homeostasis, modulated by social and exploratory motives. At that point, we will be able to say we truly understand the process of cognitive development.

[1] This section is based on a study of enaction as a framework for development in cognitive robotics [385].

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.

A.1.2 EPIC — Executive Process Interactive Control

EPIC [196] is a cognitive architecture that was designed to link high-fidelity models of perception and motor mechanisms with a production system. An EPIC model requires both knowledge encapsulated in production rules and perceptual-motor parameters. There are two types of parameter: standard or system parameters which are fixed for all tasks (*e.g.* the duration of a production cycle in the cognitive processor: 50 ms) and typical parameters which have conventional values but can vary between tasks (*e.g.* the time required to effect recognition of shape by the visual processor: 250 ms).

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

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

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

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

A.1.3 ACT-R — Adaptive Control of Thought - Rational

The ACT-R [6, 7] cognitive architecture is a widely-regarded candidate for a unified theory of cognition. It focusses on modular decomposition and offers a theory of how these modules are integrated to produce coherent cognition. The architecture comprises five specialized modules, each devoted to processing a different kind of information (see Figure A.1). There is a vision module for determining the identity and position of objects in the visual field, a manual module for controlling hands, a declarative module for retrieving information from long-term information, and a goal module for keeping track of the internal state when solving a problem. Finally, it also has a production system that coordinates the operation of the other four modules. It does this indirectly via four buffers into which each module places a limited amount of information.

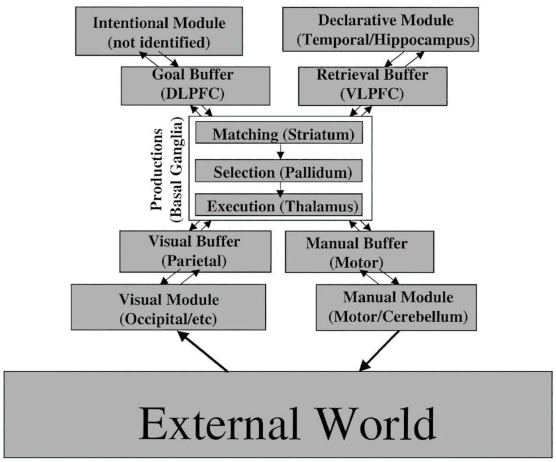


Fig. A.1 The ACT-R Cognitive Architecture (from [7])

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

There are two serial bottle-necks in ACT-R. One is that the content of any buffer is limited to a single declarative unit of knowledge, called a 'chunk'. This implies that only one memory can be retrieved at a time and indeed that a single object can be encoded in the visual field at any one time. The second bottle-neck is that only one production is selected to fire in any one cycle. This contrasts with both Soar and

EPIC both of which allow many productions to fire. When multiple production rules are capable of firing, an arbitration procedure called conflict resolution is activated.

Whilst early incarnations of ACT-R focussed primarily on the production system, the importance of perceptuo-motor processes in determining the nature of cognition is recognized by Anderson *et al.* in more recent versions [55, 7]. That said, the perceptuo-motor system in ACT-R is based on the EPIC architecture [196] which doesn't deal directly with real sensors or motors but simply models the basic timing behaviour of the perceptual and motor systems. In effect, it assumes that the perceptual system has already parsed the visual data into objects and associated sets of features for each object [6]. Anderson *et al.* recognize that this is a short-coming, remarking that ACT-R implements more a theory of visual attention than a theory of perception, but hope that the ACT-R cognitive architecture will be compatible with more complete models of perceptual and motor systems. The ACT-R visual module differs somewhat from the EPIC visual system in that it is separated into two sub-modules, each with its own buffer, one for object localization and associated with the dorsal pathway, and the other for object recognition and associated with the ventral pathway. Note that this sharp separation of function between the ventral and dorsal pathways has been challenged by recent neurophysiological evidence which points to the interdependence between the two pathways [316, 314]. When the production system requests information from the localization module, it can supply constraints in the form of attribute-value pairs (*e.g.* colour-red) and the localization module will then place a chunk in its buffer with the location of some object that satisfies those constraints. The production system queries the recognition system by placing a chunk with location information in its buffer; this causes the visual system to subsequently place a chunk representing the object at that location in its buffer for subsequent processing by the production system. This is a significant idealization of the perceptual process.

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

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

that is retrieved (*i.e.* that exceeds the threshold) is an exponential function of the activation.

Procedural memory is encapsulated in the production system which coordinates the overall operation of the architecture. Whilst several productions may qualify to fire, only one production is selected. This selection is called conflict resolution. The production selected is the one with the highest utility, a factor which is a function of an estimate of the probability that the current goal will be achieved if this production is selected, the value of the current goal, and an estimate of the cost of selecting the production (typically proportional to time), both of which are learned in a Bayesian framework from previous experience with that production. In this way, ACT-R can adapt to changing circumstances [55].

Declarative knowledge effectively encodes things in the environment while procedural knowledge encodes observed transformations; complex cognition arises from the interaction of declarative and procedural knowledge [6]. A central feature of the ACT-R cognitive architecture is that these two types of knowledge are tuned in specific application by encoding the statistics of knowledge. Thus, ACT-R learns sub-symbolic information by adjusting or tuning the knowledge parameters. This sub-symbolic learning distiguishes ACT-R from the symbolic (production-rule) learning of Soar.

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

Like Soar, ACT-R has evolved significantly over several years [6]. It is currently in Version 6.0.

A.1.4 The ICARUS Cognitive Architecture

The ICARUS cognitive architecture [60, 213, 214, 215, 216] follows in the tradition of other cognitivist architectures, such ACT-R, Soar, and EPIC, exploiting symbolic representations of knowledge, the use of pattern matching to select relevant knowledge elements, operation according to the conventional recognize-act cycle, and an incremental approach to learning. In this, ICARUS adheres strictly to Newell's and Simon's physical symbol system hypothesis [272] which states that symbolic processing is a necessary and sufficient condition for intelligent behaviour. However, ICARUS goes further, asserting that mental states are always grounded in either real or imagined physical states, and *vice versa* that problem-space symbolic operators always expand to actions that can be effected or executed. Langley refers to this as the *symbolic physical system* hypothesis. This assertion of the importance of action and perception is similar to recent claims by others in the cognitivist community such as Anderson *et al.* [7].

There are also some other important difference between ICARUS and other cognitivist architectures. ICARUS distinguishes between concepts and skills, and devotes two different types of representation and memory for them, with both long-term and short-term variants of each. Conceptual memory encodes knowledge about general classes of objects and relations among them whereas skill memory encodes knowledge about ways to act and achieve goals. ICARUS forces a strong correspondence between short-term and long-term memories, with the short-term structured being specific instances of the long-term structures. These instances are triggered on the basis of the contents of another short-term memory — a perceptual buffer — which contains a description of physical entitites that correspond to the output of sensors. Furthermore, ICARUS adopts a strongly hierarchical organization for its long-term memory, with conceptual memory directing bottom-up inference and skill memory structuring top-down selection of actions.

Langley notes that incremental learning is central to most cognitivist cognitive architectures, in which new cognitive structures are created by problem solving when an impasse is encountered. ICARUS adopts a similar stance so that when an execution module cannot find an applicable skill that is relevant to the current goal, it resolves the impasse by backward chaining. ICARUS differs from other cognitivist cognitive architectures in that it focusses on the origin of hierarchical skill-based structures which Langley suggest arise incrementally from problem-solving behaviour. Like many other cognitivist cognitive architectures such as Soar and ACT-R, ICARUS maintains a commitment to a unified theory of cognition that is consistent with human capabilities.

The ICARUS execution cycle operates as follows. First, the perceptual buffer is updated. These structures are compared with the long-term concept memory and those that match are instantiated in the short-term concept memory as beliefs. The ICARUS long-term concept memory is organized as a lattice structure, with primitive concepts at the lowest level and more complex composite concepts at successively higher levels. Equally, instances of every level of matching concept are created in the short-term memory, provided they have support in the perceptual

buffer. The short-term skill memory is then examined to determined which skill should be considered for execution, based on its current goals. To ensure that this execution does not degenerate into a simple stimulus-response behaviour, ICARUS uses a global persistence parameter to evaluate possible contenting skills for execution. The higher the persistence factor, the greater the system's bias for selecting the skills it chose on the previous cycle.

The skill memories, both short-term and long-term, are organized as hierarchical structures of primitive and non-primitive skills. Primitive skills comprise an action sequence, the concepts that must hold to initiate the skill, and a description of the resultant state should the skill be executed. Non-primitive skills specify how to decompose the skill further and a description of the concepts that will be achieved upon successful execution of the skill. A recent version of ICARUS [216] provides an extension which allows the architecture to compose skills to solve new problems and to store them, *i.e.*, to learn new skills. To date, ICARUS has not yet been used with a physical robot although it is the stated intention of Langley *et al.* to do so [216].

A.1.5 ADAPT — A Cognitive Architecture for Robotics

Some authors, e.g. Benjamin *et al.* [30], argue that existing cognitivist cognitive architectures such as Soar, ACT-R, and EPIC, don't easily support certain mainstream robotics paradigms such as adaptive dynamics and active perception. Many robot programs comprise several concurrent distributed communicating real-time behaviours and consequently these architectures are not suited since their focus is primarily on "sequential search and selection", their learning mechanisms focus on composing sequential rather than concurrent actions, and they tend to be hierarchically-organized rather than distributed. Benjamin *et al.* don't suggest that you cannot address such issues with these architectures but that they are not central features. They present a different cognitive architecture, ADAPT — Adaptive Dynamics and Active Perception for Thought, which is based on Soar but also adopts features from ACT-R (such as long-term declarative memory in which sensori-motor schemas to control perception and action are stored) and EPIC (all the perceptual processes fire in parallel) but the low-level sensory data is placed in short-term working memory where it is processed by the cognitive mechanism. ADAPT has two types of goals: task goals (such as 'find the blue object') and architecture goals (such as 'start a schema to scan the scene'). It also has two types of actions: task actions (such as 'pick up the blue object') and architectural actions (such as 'initiate a grasp schema'). While the architectural part is restricted to allow only one goal or action at any one time, the task part has no such restrictions and many task goals and actions — schemas — can be operational at the same time. The architectural goals and actions are represented procedurally (with productions) while the task goals and actions are represented declaratively in working memory as well as procedurally.

A.1.6 The GLAIR Cognitive Architecture

GLAIR (Grounded Layered Architecture with Integrated Reasoning) [339] is three-layer cognitive architecture comprising a low-level Sensori-Actuator Layer (SAL), a mid-level Perceputo-Motor Layer (PML), and a high-level Knowledge Layer (KL). The SAL controls the sensors and the hardware effectors. The PML is divided into three sublayers which are, from bottom to top, the PMLc layer which encapsulates the sensors and the effectors in the robot's repertoire of behaviours, the PMLb which acts as an interface with the PMLa, and the PMLa itself which serves to ground the symbolic knowledge in the KL in perceptions and actions. The KL represents the beliefs of the agent and it is at this layer that reasoning, planning, and act selection are effected.

GLAIR is a strongly cognitivist architecture, advocating both mind-body dualism (*i.e.* the logical separation of mind and body) and and functionalism (cognitive mechanisms are independent of the physical platform [105]), *viz.*: "The KL constitutes the mind of the agent; the PML and SAL, its body. However, the KL and PMLa layers are independent of the implementation of the agent's body, and can be connected, without modification, to a hardware robot or to a variety of software-simulated robots or avatars".

The KL is the core of the architecture insofar as it grew out of earlier work on knowledge representation and reasoning, in particular logic-based, frame-based, and network-based representations. It supports metaproposition (propositions about propositions) as well as forward and backward reasoning and bidirectional inference by treating the KL representation as a propositional graph and traversing it accordingly. According to its developers, GLAIR's focus on reasoning differentiates it from other cognitive architectures that are driven primarily by problem-solving or goal-achievement. Reflecting GLAIR's heritage as an interactive natural language understanding system, it operates using a sense-reason-act cycle whereby a natural language utterance is input, analyzed in the context of current beliefs, and a natural language utterance expressing the resultant proposition is output. The output proposition depends on whether the input utterance is a statement, a question, or a command (in which case, the output proposition will represent a new belief, an answer to the question, or an act to be performed.)

GLAIR distinguishes between different types of acts: external, internal (mental), and control acts. This act is composed of an action and zero or more arguments. An agent may perform an act, or it may have propositions about acts, or it may have a policy about an act, *i.e.* a specification of the circumstances under which reasoning leads to an act. Thus, a GLAIR agent "performs an act, believes a proposition, and adopts a policy". External acts either sense or affect the agent's environment (which can be real, virtual, or simulated). *No external acts are predefined in GLAIR and they must be supplied by the designer of the agent.* Mental acts affect the agent's beliefs and policies. Control acts determine the overall functioning of the reasoning system.

In principle, GLAIR agents are able to reason about themselves by including a term that refers to the agent itself but providing GLAIR agents with knowledge of

the actions they are currently performing above the level of primitive actions is the subject of further research. Similarly, while the KL contains delarative knowledge and procedural knowledge for carrying out pre-defined acts, GLAIR does not yet have the capability to learn these procedural representations. Finally, when attempting to achieve a goal, a GLAIR agent selects an act in the belief that it will lead to the achievment of that goal. However, GLAIR agents don't yet have the ability to formulate these beliefs by reasoning about the effects of various acts or by observing the effects of its acts. In other words, GLAIR doesn't yet learn to anticipate the outcome of its actions.

A.1.7 CoSy Architecture Schema

Sloman and his co-workers advocate splitting the design of a cognitive architecture into three steps: the analysis of several scenarios to identify the principal requirements, the creation of an architecture schema, and the instantiation of the architecture schema in a scenario-specific cognitive architecture design [137, 138]. The architecture schema addresses the organization of information and processing components and the control of information flow among them in a task and implementation independent manner. In other words, the architecture schema identifies the general shape of an architecture suitable for a relatively broad class of scenarios but leaves the details to a subsequent scenario-specific design phase.

Based on scenarios for a hypothetical robotic assistant which can interact with a family in a home environment, which can learn about and alter its world through physical actions, which can engage in linguistic discourse, and which can perform household tasks, Sloman *et al.* introduce the CoSy Architecture Schema which forms the basis of subsequent (outline) architecture instantiations [138] as well as a software toolset to effect this instantiation [137]. Because this schema is based on scenarios which focus on the robot behaviour, such a schema is neither a unified theory of cognition or general theory of mind (as many other cognitivist cognitive architectures are) nor is it a general schema which can be used to describe *any* cognitive architecture, such as Sloman's CogAff schema [347].

The CoSy Architecture Schema is based on three general requirements: that the architecture schema should support interaction in a dynamic world, that it should enable the integration of information from several sources, and that it should facilitate goal-directed behaviour with multiple goals. These requirements mean that the architecture schema must provide mechanisms for relating information from different sources, for dealing with inconsistencies, and for allowing global motives to influence processing throughout the system. Furthermore, since the target robot system is intended to have several individual capabilities or competences, the schema must allow for multiple specialized representations for each competence or sub-system. Furthermore, these must update asynchronously and concurrently in the target architecture. Sloman *et al.* claim that this focus on concurrent modular processing distinguishes the CoSy Architecture Schema from many other existing architectures which are "essentially monolithic in their styles of processing" [138].

Based on these requirements, the CoSy Architecture Schema comprises a separate subarchitecture for each competence. Each subarchitecture is derived from a common template (to be described below) and all subarchitectures are loosely-coupled to avoid complex interdependencies. Typically, there are subarchitectures for motor control, vision, planning, linguistic communication, spatio-temporal memory, and so on. Each subarchitecture comprises a set of processing components connected to a local subarchitecture-specific working memory which can be written to only by the subarchitecture processes and by a single global process (the goal manager).

The information in each subarchitecture and in the architecture schema as a whole is controlled by goals. These goals are generated by the system as it

operates and interacts with its environment. There are two types of goal. Global goals, which require coordination across two or more subarchitectures, and local goals, which originate in and are particular to a given subarchitecture.

The knowledge encapsulated in each subarchitecture is defined by a subarchitecture-specific ontology: "a structured description of the kinds of information that the cognitive system can process from either internal or external sources" [138]. Relationships between entities in these ontologies are defined by a set of general ontologies. These general ontologies define knowledge at global level. They deal with knowledge that is independent of competence and related to general architecture issues such as global goals, planning, and reasoning. The implementation of these ontologies is not specified.

Since the CoSy Architecture Schema comprises many concurrent processing components, the issue of coordination and control is an important issue. In the CoSy Architecture Schema, control is effected by the global goal manager. It is achieved by requiring each component to announce its intention to perform some processing by posting a goal for approval. Each subarchitecture has a task manager process with determines whether or not that local goal should be adopted. This decision can also be made by the global goal manager based on the current global goals for the entire system.

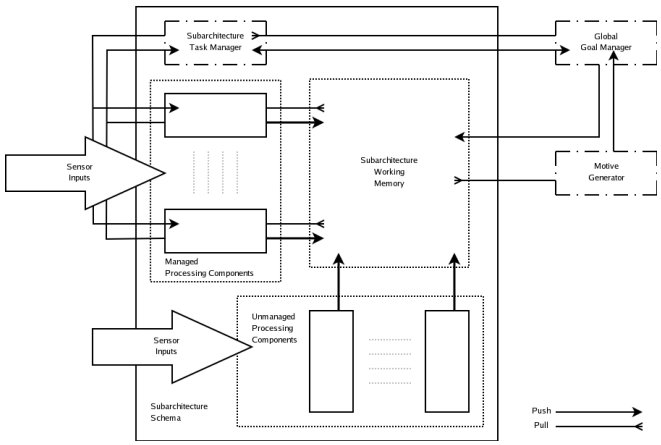


Fig. A.2 The CoSy Architecture Schema: template subarchitecture (from [138])

A template for a subarchitecture is shown in Figure A.2. Each subarchitecture comprises the following four components:

1. *Subarchitecture Working Memory* which holds the results of the processes in that subarchitecture;
2. *Subarchitecture Task Manager* which decided which of the subarchitecture's local goals is to be adopted;
3. A set of *Unmanaged Processing Components* which typically serve the systems sensor inputs and which can run *without* posting a goal for approval;
4. A set of *Managed Processing Components* which monitor the working memory for information that they can process and which post a local goal when they can process that information.

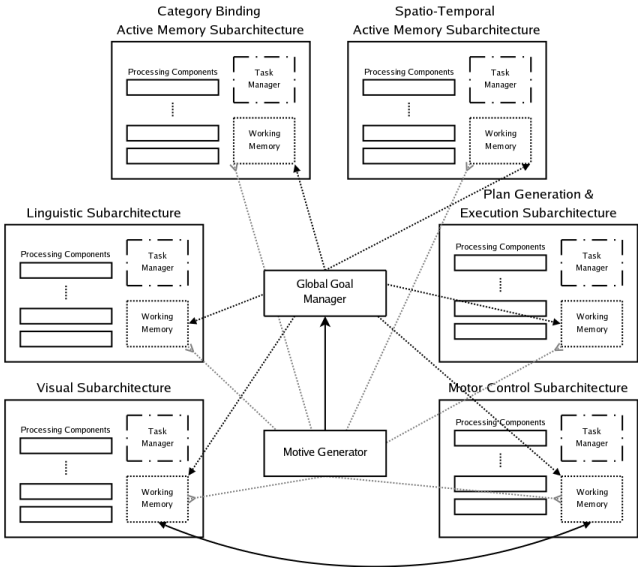


Fig. A.3 The CoSy Architecture Schema: Example architecture instatiation with six subarchitectures, the Global Goal Manager, and the Motive Generator (from [138]).

Apart from a specific subarchitecture for each competence, there are three global components: the *Motive Generator*, the *Global Goal Manager*, and the *General Memory* (see Figure A.3). The motive generator monitors the working memory in every subarchitecture looking for information which may be able to trigger processing in a different subarchitecture. When it finds such information, it posts a global goal for the whole architecture. This global goal is then considered by the global goal manager (the manner in which the global goals are adopted is not specified). Once a global goal is adopted, it writes information to the working memory of the appropriate subarchitecture which will in turn cause a component in the subarchitecture to post a local goal and start a chain of local processing. The general memory

stores long-term knowledge about anything that is relevant to the control of the overall instantiated architecture schema, *e.g.* beliefs, goals, etc. Again, no specific structure for this general memory is proposed in [138].

Ultimately, the CoSy Architecture Schema is a cognitivist rule-based schema for the design of cognitive architectures focussing on the effective coordination of multiple concurrent asynchronously-updating sub-systems, achieved through a combination of local and global moderation of goal-oriented processing and through the sharing and integration of knowledge among these sub-systems. The manner in which these goals are acquired or how knowledge is generated is not specified. Learning is not specifically addressed in the schema although it is evident that a capacity for learning is envisaged to be incorporated in some of the processing components in one or more of the competence subarchitectures.

A.2 Emergent Cognitive Architectures

A.2.1 Autonomous Agent Robotics

Autonomous agent robotics (AAR) and behaviour-based systems represents an emergent alternative to cognitivist approaches. Instead of a cognitive system architecture that is based on a decomposition into functional components (*e.g.* representation, concept formation, reasoning), an AAR architecture is based on interacting *whole* systems. Beginning with simple whole systems that can act effectively in simple circumstances, layers of more sophisticated systems are added incrementally, each layer subsuming the layers beneath it. This is the subsumption architecture introduced by Brooks [49]. Christensen and Hooker [62] argue that AAR is not sufficient either as a principled foundation for a general theory of situated cognition. One limitation includes the explosion of systems states that results from the incremental integration of sub-systems and the consequent difficulty in coming up with an initial well-tuned design to produce coordinated activity. This in turn imposes a need for some form of self-management, something not included in the scope of the original subsumption architecture. A second limitation is that it becomes increasingly problematic to rely on environmental cues to achieve the right sequence of actions or activities as the complexity of the task rises. AAR is also insufficient for the creation of a comprehensive theory of cognition: as the subsumption architecture can't be scaled to provide higher-order cognitive faculties (it can't explain self-directed behaviour) and even though the behaviour of an AAR system may be very complex it is still ultimately a reactive system.

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

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

A.2.2 A Global Workspace Cognitive Architecture

Shanahan [335, 336, 337, 338] proposes a biologically-plausible brain-inspired neural-level cognitive architecture in which cognitive functions such as anticipation and planning are realized through internal simulation of interaction with the environment. Action selection, both actual and internally simulated, is mediated by affect. The architecture is based on an external sensori-motor loop and an internal sensori-motor loop in which information passes though multiple competing cortical areas and a global workspace.

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

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

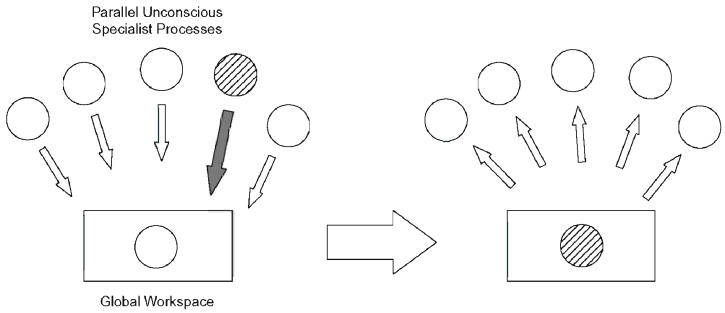


Fig. A.4 The Global Workspace Theory cognitive architecture: 'winner-take-all' coordination of competing concurrent processes (from [337])

Shanahan's cognitive architecture is founded also upon the fundamental importance of parallelism as a constituent component in the cognitive process as opposed to being a mere implementation issue. He deploys the *global workspace* model [15, 16] of information flow in which a sequence of states emerges from the interaction of many separate parallel processes (see Figure A.4). These specialist processes compete and co-operate for access to a global workspace. The winner(s) of

the competition gain(s) controlling access to the global access and can then broadcast information back to the competing specialist processes. Shanahan argues that this type of architecture provides an elegant solution to the frame problem.

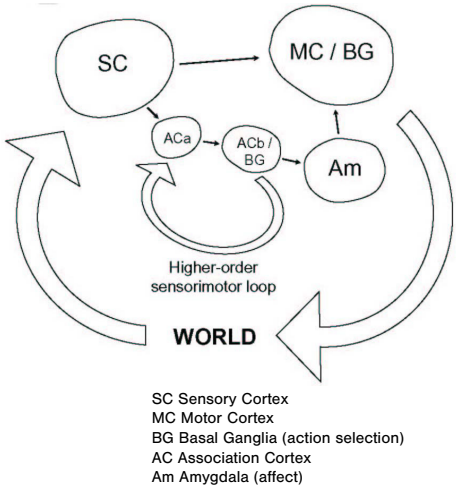


Fig. A.5 The Global Workspace Theory cognitive architecture: achieving prospection by sensori-motor simulation (from [337])

Shanahan's cognitive architecture is comprised of the following components: a first-order sensori-motor loop, closed externally through the world, and a higher-order sensori-motor loop, closed internally through associative memories (see Figure A.4). The first-order loop comprises the sensory cortex and the basal ganglia (controlling the motor cortex), together providing a reactive action-selection sub-system. The second-order loop comprises two associative cortex elements which carry out off-line simulations of the system's sensory and motor behaviour, respectively. The first associative cortex simulates a motor output while the second simulates the sensory stimulus expected to follow from a given motor output. The higher-order loop effectively modulates basal ganglia action selection in the first-order loop via an affect-driven amygdala component. Thus, this cognitive architecture is able to anticipate and plan for potential behaviour through the exercise of its "imagination" (*i.e.* its associative internal sensori-motor simulation). The global workspace doesn't correspond to any particular localized cortical area. Rather, it is a global communications network.

The architecture is implemented as a connectionist system using G-RAMs: generalized random access memories [3]. Interpreting its operation in a dynamical framework, the global workspace and competing cortical assemblies each define

an attractor landscape. The perceptual categories constitute attactors in a state space that reflects the structure of the raw sensory data. Prediction is achieved by allowing the higher-order sensori-motor loop to traverse along a simulated trajectory in that state space so that the global workspace visits a sequence of attractors. The system has been validated in a Webot [252] simulation environment.

A.2.3 Self-directed Anticipative Learning

Christensen and Hooker propose a new emergent interactivist-constructivist (I-C) approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [61]. This approach falls under the broad heading of dynamical embodied approaches in the non-cognitivist paradigm. They assert first the primary model for cognitive learning is anticipative skill construction and that processes that both guide action and improve the capacity to guide action while doing so are taken to be the root capacity for all intelligent systems. For them, intelligence is a continuous management process that has to support the need to achieve autonomy in a living agent, distributed dynamical organization, and the need to produce functionally coherent activity complexes that match the constraints of autonomy with the appropriate organization of the environment across space and time through interaction. In presenting their approach they use the term "explicit norm signals" for the signals that a system uses to differentiate an appropriate context performing an action. These norm signals reflect conditions for the (maintenance) of the system's autonomy (*e.g.* hunger signals depleted nutritional levels). The complete set of norm signals is termed the norm matrix. They then distinguish between two levels of management: low-order and high-order. Low-order management employs norm signals which differentiate only a narrow band of the overall interaction process of the system (*e.g.* a mosquito uses heat tracking and CO_2 gradient tracking to seek blood hosts). Since it uses only a small number of parameters to direct action, success ultimately depends on simple regularity in the environment. These parameters also tend to be localized in time and space. On the other hand, high-order management strategies still depend to an extent on regularity in the environment but exploit parameters that are more extended in time and space and use more aspects of the interactive process, including the capacity to anticipate and evaluate the system's performance, to produce effective action (and improve performance). This is the essence of self-directedness. "Self-directed systems anticipate and evaluate the interaction process and modulate system action accordingly". The major features of self-directedness are action modulation ("generating the right kind of extended interaction sequences"), anticipation ("who will/should the interaction go?", evaluation ("how did the evaluation go?"), and constructive gradient tracking ("learning to improve performance").

A.2.4 A Self-Affecting Self-Aware (SASE) Cognitive Architecture

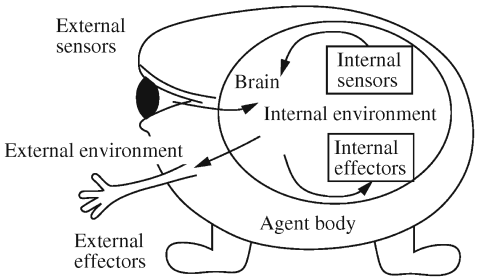


Fig. A.6 The Self-Aware Self-Effecting (SASE) architecture (from [393])

Weng [394, 395, 393] introduced an emergent cognitive architecture that is specifically focussed on the issue of development, by which he means that the processing accomplished by the architecture is not specified (or programmed) *a priori* but is the result of the real-time interaction of the system with the environment including humans. Thus, the architecture is not specific to tasks, which are unknown when the architecture is created or programmed, but is capable of adapting and developing to learn both the tasks required of it and the manner in which to achieve the tasks. In this sense, even through Weng's architecture is not a cognitivist one, his use of the term is very faithful to the meaning of *cognitive architecture* as it was originally intended when it was introduced originally in the cognitivist paradigm. That is, it represents the underlying infrastructure for a cognitive system, specifically those aspects of a cognitive agent that are constant over time and independent of the task [128, 214, 309].

Weng refers to his architecture as a Self-Aware Self-Effecting (SASE) system (see Figure A.6). The architecture entails an important distinction between the sensors and effectors that are associated with the environment (including the system's body and thereby including proprioceptive sensing) and those that are associated with the system's 'brain' or central nervous system (CNS). Only those systems that have explicit mechanisms for sensing and affecting the CNS qualify as SASE architectures. The implications for development are significant: the SASE architecture is configured with no knowledge of the tasks it will ultimately have to perform, its brain or CNS are not directly accessible to the (human) designers once it is launched, and after that the only way a human can affect the agent is through the external sensors and effectors. Thus, the SASE architecture is very faithful to the emergent paradigms of cognition, especially the enactive approach: its phylogeny is fixed and it is only through ontogenetic development that the system can learn to operate effectively in its environment.

The concept of self-aware self-effecting operation is similar to the level 2 autopoietic organizational principles introduced by Matura and Varela [237] (*i.e.* both self-production and self-development) and is reminiscent of the recursive self-maintenant systems principles of Bickhard [40] and Christensen's and Hooker's interactivist-constructivist approach to modelling intelligence and learning: self-directed anticipative learning (SDAL) [61]. Weng's contribution differs in that he provides a specific computational framework in which to implement the architecture. Weng's cognitive architecture is based on Markov Decision Processes (MDP), specifically a developmental observation-driven self-aware self-effecting Markov Decision Process (DOSASE MDP). Weng places this particular architecture in a spectrum of MDPs of varying degrees of behavioural and cognitive complexity [395]; the DOSASE MDP is type 5 of six different types of architecture and is the first type in the spectrum that provides for a developmental capacity. Type 6 builds on this to provide additional attributes, specifically greater abstraction, self-generated contexts, and a higher degree of sensory integration.

The example DOSASE MDP vision system detailed in [394] further elaborates on the cognitive architecture, detailing three types of mapping in the information flow within the architecture: sensory mapping, cognitive mapping, and motor mapping. It is significant that there is more than one cognitive pathway between the sensory mapping and the motor mapping, one of which encapsulates innate behaviours (and the phylogenetically-endowed capabilities of the system) while the other encapsulates learned behaviours (and the ontogenetically-developed capabilities of the system). These two pathways are mediated by a subsumption-based motor mapping which accords higher priority to the ontogenetically-developed pathway. A second significant feature of the architecture is that it facilitates what Weng refers to as "primed sensations" and "primed action". These correspond to predictive sensations and actions and thereby provide the system with the anticipative and prospective capabilities that are the hallmark of cognition.

The general SASE schema, including the associated concept of Autonomous Mental Development (AMD), has been developed and validated in the context of two autonomous developmental robotics systems, SAIL and DAV [396, 397, 394, 395].

A.2.5 Darwin: Neuromimetic Robotic Brain-Based Devices

Kirchmar *et al.* [203, 204, 205, 206, 207, 334] have developed a series of robot platforms called Darwin to experiment with developmental agents. These systems are 'brain-based devices' (BBDs) which exploit a simulated nervous system that can develop spatial and episodic memory as well as recognition capabilities through autonomous experiential learning. As such, BDDs are a neuromimetic approach in the emergent paradigm that is most closely aligned with the enactive and the connectionist models. It differs from most connectionist approaches in that the architecture is much more strongly modelled on the structure and organization of the brain than are conventional artificial neural networks, *i.e.* they focus on the nervous system as a whole, its constituent parts, and their interaction, rather than on a neural implementation of some individual memory, control, or recognition function.

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

Darwin VIII is capable of discriminating reasonably simple visual targets (coloured geometric shapes) by associating it with an innately preferred auditory cue. Its simulated nervous system contains 28 neural areas, approximately 54,000 neuronal units, and approximately 1.7 million synaptic connections. The architecture comprises regions for vision (V1, V2, V4, IT), tracking (C), value or saliency (S), and audition (A). Gabor filtered images, with vertical, horizontal, and diagonal selectivity, and red-green colour filters with on-centre off-surround and off-centre on-surround receptive fields, are fed to V1. Sub-regions of V1 project topographically to V2 which in turn projects to V4. Both V2 and V4 have excitatory and inhibitory reentrant connections. V4 also has a non-topographical projection back to V2 as well as a non-topographical projection to IT, which itself has reentrant adaptive connections. IT also projects non-topographically back to V4. The tracking area (C) determines the gaze direction of Darwin VIII's camera based on excitatory projections from the auditory region A. This causes Darwin to orient toward a sound source. V4 also projects topographically to C causing Darwin VIII to centre its gaze on a visual object. Both IT and the value system S have adaptive connections to C which facilitates the learned target selection. Adaptation is effected using the Hebbian-like Bienenstock-Cooper-Munroe (BCM) rule [41]. From a behavioural perspective, Darwin VIII is conditioned to prefer one target over others by associating it with the innately preferred auditory cue and to demonstrate this preference by orienting towards the target.

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

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

A.2.6 The Cognitive-Affective Architecture

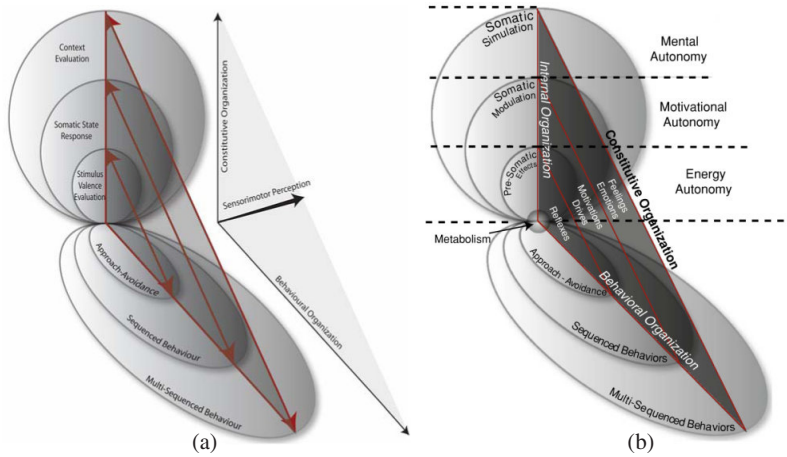


Fig. A.7 The cognitive-affective enactive architecture: (a) The Enactive Organizational Constraints Hierarchy (from [264]) in which increasing constitutive organizational complexity facilitates increasing levels of stimulus evaluation and appraisal in maintaining the constitutive autonomy of the system, accompanied by an associated increase in adaptivity. These levels are directly coupled, by sensorimotor perception, to increasing complexity along the behavioural organization axis; (b) The Cognitive-Affective Architecture Schematic (from [415]) refines this by showing a single spectrum of constitutive organization brought about by the recruitment of a progression of emotions, from reflexes, through drives and motivations, to emotions-proper and feelings. Each level in the constitutive organization is associated on the Internal Organization axis with an increasing level of homeostatic autonomy-preserving self-maintenance, ranging from basic metabolic processes through reactive sensorimotor activity (pre-somatic effects), associative learning and prediction (somatic modulation), to interoception and internal simulation of behaviour prior to action. Equally, each level in the constitutive organization is associated on the Behavioural Organization axis with an increasing level of complexity in behaviour, ranging from approach-avoidance, sequenced behaviours, and multi-sequenced behaviours.

Ziemke and his co-workers have developed a schema for an enactive cognitive architecture [264, 415] that explicitly addresses the role of emotion in a cognitive system. Based on the architecture and physiology of the mammalian brain, they refer to it both as a schema for an "Enactive Organizational Constraints Hierarchy" [264] and a "Cognitive-Affective Architecture Schematic" [415]. It is a schema in the sense that it identifies the principal characteristics of the architecture without

providing a detailed design of the component parts of the architecture and the dynamics of their interaction. However, to complement the schema, they also propose a design process called *holistic-reductionism* which focusses on the interdependencies of the components rather than on the identification of independent functional modules, as is normally the case with architecture design. Any modularity in the system, they argue, emerges from the interdependence of the embodied cognitive processes rather than by phylogenetic pre-specificiation.

The key idea in holistic reductionism is to consider first a minimally-cognitive agent which is impaired by de-cortication of selected cognitive functions and then to incrementally re-corticate the agent, allowing progressively greater cognitive capabilities to emerge. Thus, the methodology is to realize a minimal, but complete and viable, autonomous system and then develop it through re-cortication. Targetting an initial de-corticated system allows one to focus on the essential requirements of autonomous self-maintenance through essential metabolic homeostatic processes. These processes can be perturbed when interacting with the world in which the system is embedded and "the well-being of the agent, specified in terms of disruption and the effort required to re-assert metabolic norms provides the basis of motivation" [264]. The autonomy of the agent is effected through a hierarchy of homeostatic self-regulatory processes, each of which exploits a progression of associated emotions, ranging from basic reflexes linked to metabolic regulation, through drives and motivations, to emotions-proper and feelings linked to "higher" cognitive functions. This follows closely Damasio's hierarchy of levels of homeostatic regulation [73]. Thus, "the emotional aspect of cognition, providing motivation and value to an otherwise neutral world, ... is a fundamental part of the make-up of and organism with respect to sensorimotor learning" [264]. When extending the autonomy — and cognitive capabilities — of the system by re-cortication, Ziemke emphasizes that "new elements of the extended model must integrate with the existing model through modulation and not by functional replacement, or by modular extension" [264].

The Enactive Organizational Constraints Hierarchy (see Fig. A.7 (a)) traverses two dimensions: (i) Constitutive Organization and (ii) Behavioural Organization. The former refers to the system's internal dynamics as it maintains its integrity — its autonomy — in the face of perturbation by various stimuli. At the core of this space there is metabolic homeostatic self-regulation. This extends, as re-cortication proceeds, to stimulus valence evaluation, somatic state response, and content evaluation, each level offering increasing organizational complexity, an increasing degree of decoupling between stimulus and response, and an increasing degree of appraisal and associated adaptivity. Each level in the constitutive organization dimension is matched by an associated level in the behavioural organization dimension: approach-avoidance, sequenced behaviour, and multi-sequenced behaviour, respectively. Thus, the behavioural organization dimension is coupled by sensorimotor perception to the constitutive organizational dimension. A later version of the architecture [415], the Cognitive-Affective Architecture Schematic, reflects this coupling by referring to a single space of constitutive organization which is viewed from two perspectives: internal organization and behavioural organization (see Fig. A.7 (b)).

The spectrum of constitutive organization is realized by the recruitment of a progression of emotions, from reflexes, through drives and motivations, to emotions-proper and feelings. Each level in the constitutive organization is associated on the Internal Organization axis with an increasing level of homeostatic autonomy-preserving self-maintenance, ranging from basic metabolic processes through reactive sensorimotor activity (pre-somatic effects), associative learning and prediction (somatic modulation), to interoception and internal simulation of behaviour prior to action. Equally, each level in the constitutive organization is associated on the Behavioural Organization axis with an increasing level of complexity in behaviour, ranging from approach-avoidance, sequenced behaviours, and multi-sequenced behaviours.

The key idea under-pinning the Cognitive-Affective Architecture is that different levels of cognitive function and behavioural complexity are associated with, and are brought about by, different levels of emotion, each linked to affective homeostatic processes ranging from reflexes right through to internal simulation.

A.3 Hybrid Cognitive Architectures

A.3.1 A Humanoid Robot Cognitive Architecture

Burghart *et al.* [54] present a hybrid cognitive architecture for a humanoid robot. It is based on interacting parallel behaviour-based components, comprising a three-level hierarchical perception sub-system, a three-level hierarchical task handling system, a long-term memory sub-system based on a global knowledge database (utilizing a variety of representational schemas, including object ontologies and geometric models, Hidden Markov Models, and kinematic models), a dialogue manager which mediates between perception and task planning, an execution supervisor, and an 'active models' short-term memory sub-system to which all levels of perception and task management have access. These active models play a central role in the cognitive architecture: they are initialized by the global knowledge database and updated by the perceptual sub-system and can be autonomously actualized and reorganized. The perception sub-system comprises a three-level hierarchy with low, mid, and high level perception modules. The low-level perception module provides sensor data interpretation without accessing the central system knowledge database, typically to provide reflex-like low-level robot control. It communicates with both the mid-level perception module and the task execution module. The mid-level perception module provides a variety of recognition components and communicates with both the system knowledge database (long-term memory) as well as the active models (short-term memory). The high-level perception module provides more sophisticated interpretation facilities such as situation recognition, gesture interpretation, movement interpretation, and intention prediction.

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

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

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

A.3.2 The Cerebus Architecture

Horswill [170, 171] argues that classical artificial intelligence systems such as those in the tradition of Soar, ART-R, and EPIC, are not well suited for use with robots. Traditional systems typically store all knowledge centrally in a symbolic database of logical assertions and reasoning is concerned mainly with searching and sequentially updating that database. However, robots are distributed systems with multiple sensory, reasoning, and motor control processes all running in parallel and often only loosely coupled with one another. Each of these processes maintains its own separate and limited representation of the world and the task at hand and he argues that it is not realistic to require them to constantly synchronize with a central knowledge base.

Recently, much the same argument has been made by neuroscientists about the structure and operation of the brain. For example, evidence suggest that space perception is not the result of a single circuit, and in fact derives from the joint activity of several fronto-parietal circuits, each of which encodes the spatial location and transforms it into a potential action in a distinct and motor-specific manner [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.

Horswill contends also that the classical reasoning systems don't have any good way of directing perceptual attention: they either assume that all the relevant information is already stored in the database or they provide a set of actions that fire task-specific perceptual operators to update specific parts of the database (just as, for example, happens in ACT-R). Both of these approaches are problematic: the former fall foul of the frame problem (the need to differentiate the significant in a very large data-set and then generalize to accommodate new data) and the second requires that the programmer design the rule based to ensure that the appropriate actions are fired in the right circumstances and at the right time; see also similar arguments by Christensen and Hooker [62].

Horswill argues that keeping all of the distinct models or representations in the distributed processes or sub-systems consistent needs to be a key focus of the overall architecture and that is should be done without sychronizing with a central knowledge base. They propose a hybrid cognitive architecture, Cerebus, that combines the tenets of behaviour-based architectures with some features of symbolic AI (forward- and backward-chaining inference using predicate logic). It represents an attempt to scale behaviour-based robots (*e.g.* see Brooks [49] and Arkin [11]) without resorting to a traditional central planning system. It combines a set of behaviour-based sensory-motor systems with a marker-passing semantic network and an inference network. The semantic network effects long-term declarative memory, providing

reflective knowledge about its own capabilities, and the inference network allows it to reason about its current state and control processes. Together they implement the key feature of the Cerebus architecture: the use of reflective knowledge about its perceptual-motor systems to perform limited reasoning about its own capabilities.

A.3.3 Cog: Theory of Mind

Cog [51] is an upper-torso humanoid robot platform for research on developmental robotics. Cog has a pair of six degree-of-freedom arms, a three degree-of-freedom torso, and a seven degree-of-freedom head and neck. It has a narrow and wide angle binocular vision system (comprising four colour cameras), an auditory system with two microphones, a three-degree of freedom vestibular system, and a range of haptic sensors.

As part of this project, Scassellati has put forward a proposal for a Theory of Mind for Cog [333] that focusses on social interaction as a key aspect of cognitive function in that social skills require the attribution of beliefs, goals, and desires to other people.

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

Scassellati's proposed architecture is based on Leslie's model of Theory of Mind [221] and Baron-Cohen's model of Theory of Mind [25] both of which decompose the problem into sets of precursor skills and developmental modules, albeit in a different manner. Leslie's Theory of Mind emphasizes independent domain specific modules to distinguish (a) mechanical agency, (b) actional agency, and (c) attitudinal agency; roughly speaking the behaviour of inanimate objects, the behaviour of animate objects, and the beliefs and intentions of animate objects. Baron-Cohen's Theory of Mind comprises three four modules, one of which is concerned with the interpretation of perceptual stimuli (visual, auditory, and tactile) associated with self-propelled motion, and one of which is concerned with the interpretation of visual stimuli associated with eye-like shapes. Both of these feed a shared attention module which in turn feed a Theory of Mind module that represents intentional knowledge or 'epistemic mental states' of other agents.

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

A.3.4 Kismet

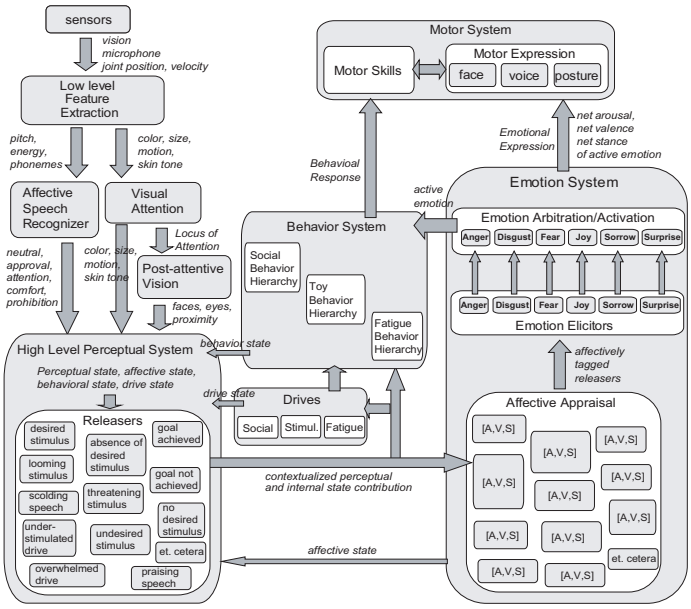


Fig. A.8 The Kismet cognitive architecture (from [46])

The role of emotion and expressive behaviour in regulating social interaction between humans and robots has been examined by Breazeal using an expressive articulated anthropomorphic robotic head called Kismet [45, 46]. Kismet has a total of 21 degree-of-freedom, three to control the head orientation, three to direct the gaze, and fifteen to control the robots facial features (*e.g.* eye-lids, eyebrows, lips, and ears). Kismet has a narrow and wide angle binocular vision system (comprising four colour cameras), and two microphones, one mounted in each ear. Kismet is designed to engage people in natural and expressive face-to-face interaction, perceiving a natural social cues and responding through gaze direction, facial expression, body posture, and vocal babbling.

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

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

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

The perceptual system comprises a set of low-level processes which sense visual and auditory stimuli, perform feature extraction (*e.g.* colour, motion, frequency), extract affective descriptions from speech, orient visual attention, and localize relevant features such as faces, eyes, objects, *etc.*. These are input to a high level perceptual system where, together with affective input from the emotion system, input from the drive system and the behaviour system, they are bound by *releaser* processes 'that encode the robot's current set of beliefs about the state of the robot and its relation to the world. There are many different kinds of releasers, each of which is 'hand-crafted' by the system designer. When the activation level of a releaser exceeds a given threshold (based on the perceptual, affective, drive, and behavioural inputs) it is output to the emotion system for appraisal. Breazeal says that 'each releaser can be thought of as a simple "cognitive" assessment that combines lower-level perceptual features with measures of its internal state into behaviorally significant perceptual categories' [46]. The appraisal process tags the releaser output with pre-specified (*i.e.* designed-in) affective information on their arousal (how much it stimulates the system), valence (how much it is favoured), and stance (how approachable it is). These are then filtered by 'emotion elicitor' to map each AVS (arousal, valence, stance) triple onto the individual emotions. A single emotion is then selected by a winner-take-all arbitration process, and output to the behaviour system and the motor system to evoke the appropriate expression and posture.

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

A.3.5 The LIDA Cognitive Architecture

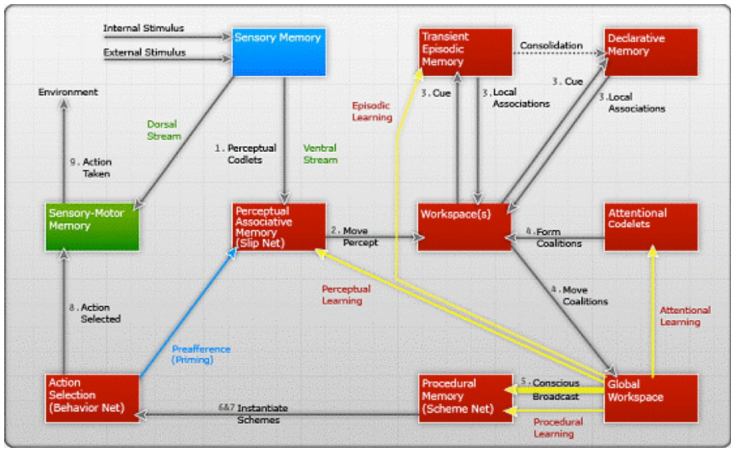


Fig. A.9 The LIDA cognitive cycle (from [17])

LIDA (Learning Intelligent Distribution Agent) is a hybrid cognitive architecture which combines features of both symbolic cognitivist and connectionist approaches [17, 103, 104, 106, 299]. It deploys several modules and processes to effect attention, action selection, and learning. The operation of LIDA is based around the concept of an atomic cognitive action-perception cycle. Each cycle comprises three phases: understanding, attending, and action selection (see Figure A.9).

The understanding phase involves sampling or sensing the environment and then it "makes sense" of its current situation by updating its representation of external sensory-derived features and internally-generated features of the agent's world comprising objects, categories, relations, events, and situations. These features are stored in a sensory memory module and a perceptual memory module, respectively.

The attending phase decides what aspect of the current situation model requires attention. This attentional process uses a mechanism adapted from Global Workspace Theory [15, 16] whereby each portion of the model competes for attention by being moved to a global workspace where a single portion of this model is selected. This portion is then broadcast back to the rest of the system. The contents of the broadcast yields a set of potential actions which are then subjected to a further competition in the subsequent action selection phase.

The initial representation of the current situation resulting from the understanding phase is stored in the perceptual memory. This is used by the workspace module to access transient and declarative episodic memories of events. Both episodic memories use these inputs to recall associatively past experiences. These

recalled perceptual events are re-assembled with the current percept and past percepts in the global workspace to generate a new model. This completes the understanding phase.

Portions of this model then compete for attention in a Global Workspace Theory winner-take-all competition. The winning portion is then broadcast globally to all other modules in the architecture. This completes the attending phase of the cycle.

The primary recipient of the broadcast is a procedural memory module which stores templates of possible associated actions and outcomes, and a measure of the likelihood of the outcome occuring. Templates that match best with the broadcast data are then passed together with templates from previous cycles to the action selection mechanism which chooses a single action for execution. The algorithms for implementing the selected action are stored in a sensorimotor memory.

LIDA uses computational versions of feelings and emotions (*i.e.* feelings with cognitive content) to modulate the operation of the action selection, attention, and learning. A representation of particular feelings are incorporated into the perceptual memory and are associated with the object representation. These are propagated though the system as part of the understanding, attending, and actions selection processes.

Learning in LIDA takes two forms: *instructionalist* (whereby new experiences are incorporated into the LIDA representation) and *selectionist* (whereby existing experiences are reinforced in the LIDA representation). Learning impacts on the three primary memories in LIDA: the perceptual memory, the episodic memory, and the procedural memory.

LIDA has only been partially implemented and, in particular, the learning aspects have not yet been completed.

A.3.6 The CLARION Cognitive Architecture

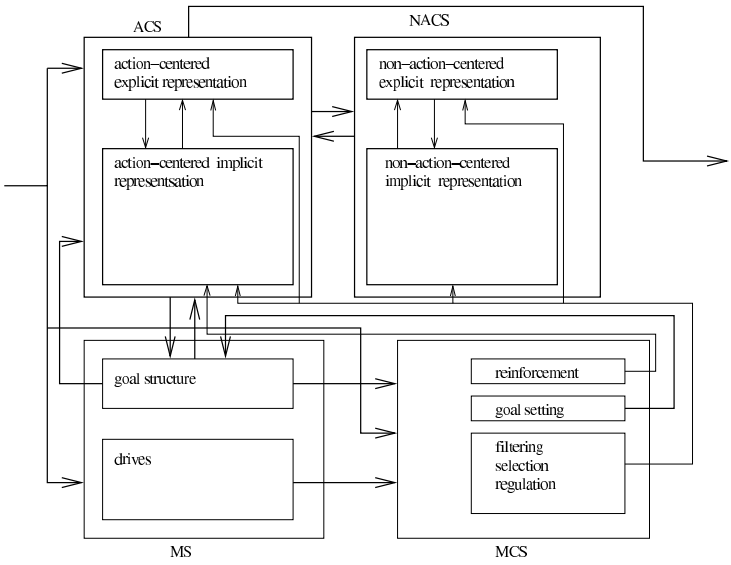


Fig. A.10 The CLARION hybrid cognitive architecture (from [364]). ACS stand for the action-centered subsystem, NACS for the non-action-centred subsystem, MS for the motivational subsystem, and MCS for the meta-cognitive subsystem. All four subsystems have two types of representation: implicit (connectionist) and explicit (symbolic).

CLARION [362, 363, 364] is an architypal hybrid cognitive architecture, deploying both connectionist and symbolic representations. It comprises four subsystems:

1. An action-centred subsystem (ACS);
2. A non-action-centred subsystem (NACS);
3. A motivational subsystem (MS);
4. A meta-cognitive subsystem (MCS).

All four subsystems have two levels of knowledge representation: an implicit connectionist bottom level and an explicit symbolic top level. The implicit and explicit levels interact and cooperate both in action selection and in learning.

The action-centred subsystem controls both external physical movements and internal "mental" operations. Given some observational state, i.e. a set of sensory features, the bottom level evaluates the desirability of all possible actions. The desirability is learned by reinforcement learning using the Q-Learning algorithm [392]. At the same time, the top level identifies possible actions from a rule network, again based on the observed sensory features. The bottom-level and top-level action are compared and the most appropriate top-level action is selected and executed. The

outcome of the action is observed and the associated sensory features are used in the bottom-level reinforcement learning process. The top-level rules are also updated on the basis of the action outcome. The bottom level comprises several modules of small neural networks, each adapted to a distinct sensory modality or task. These modules can be developed by the system based on experience (*i.e. though ontogenesis*) or they can be specified *a priori* and hard-wired into the cognitive architecture (*i.e. as the system phylogeny*). The implicit bottom level and the explicit top level representations interact to effect bottom-up learning. This operates as follows. If an action selected by the bottom level is successful, then the system extracts an explicit rule that corresponds to the sensory features and the selected action, and adds the rule to its top level rule network. Subsequently, the system verifies the extracted rule and, depending on whether the outcome is successful or unsuccessful, the rule is either generalized (made more universal and applicable to other situations) or refined (made more specific and exclusive of the current situation), respectively. In this way, the CLARION cognitive architecture is able to effect autonomous generation of explicit conceptual structures by exploiting implicit knowledge acquired by trial-and-error learning. CLARION can also effect top-down learning by integrating externally-provided knowledge in the form of explicit rule-based conceptual structures and assimilating these into the bottom level implicit representation.

The non-action-centred subsystem maintains the system's general knowledge, again both in implicit connectionist form and explicit symbolic form. The implicit bottom level uses associative memory networks whereas the explict top level encodes knowledge as a network of nodes, each node corresponding to an entity-specific chunk comprising an entity identifier (*e.g. table*) and a vector of feature dimensions / feature value pairs (*e.g.* (*size*, *large*) ... (*colour*, *white*)). The feature values are represented by nodes in the bottom level associative memory. Chunks are linked associatively. Like the action-centred subsystem, both bottom-up and top-down learning can take place in the non-action-centred subsystem.

The motivational subsystem provides the drive and feedback signals that influence the system's perceptions and actions. It provides the action-centered subsystem with goals derived from low-level drives concering physiological needs (*e.g.* need to avoid boredom) and high-level drives (*e.g.* desire for imitation of other people) which can be either primary hard-wired or secondary derived drives.

The meta-cognitive subsystem monitors and governs the overall behaviour of the cognitive system to improve cognitive performance, *e.g.* by setting goals and by setting essential parameter values.

A.3.7 The PACO-PLUS Cognitive Architecture

PACO-PLUS [201] is three-level hybrid cognitive architecture for a six-degree-of-freedom robot manipulator. The architecture comprises a low-level sensorimotor robot-vision layer, a mid-level memory layer, and high-level symbolic planning layer. The system learns object-action associations by exploring its environment. These representations are used to plan and execute sequences of actions. Unexpected errors that occur during exploration or plan execution are used to improve future performance after taking corrective action at the appropriate level, *e.g.* withdrawing the end-effector at the sensorimotor level, re-inspection of the objects in the environment at the memory level, or plan reformulation at the planning level.

The sensorimotor level is responsible for low-level robot control, camera control, and the acquisition of visual and haptic perceptual data. Visual information from a high-resolution binocular stereo rig is encapsulated at this level in a number of representations, the richest being a 3D contours. This contour data is used to identify an appropriate grasping strategy for a two-finger gripper based on pre-defined associations between grasp configurations and parts of an object. Haptic data is captured from a torque sensor on the wrist of the gripper. In its present state of development, the PACO-PLUS architecture uses a limited repertoire of object and grasp configuration representations based primarily on 3-D eliptical contours.

The memory level provides a co-joint Object Action Complexes (OACs) representation. In essence, an OAC implmements a form of affordance [118] whereby the object and the actions that the object affords the robot in terms of its ability to manipulate it are represented as a single entity. These affordances are learned autonomously by the robot. The initial grasp affordances that are hardwired in the sensorimotor level are elaborated by active exploration of the object by poking, grasping, or re-orienting it in the robot gripper. This yields a full three-dimensional visual representation of the object shape. Knowledge of the motion of the robot's arm during this exploratory phase is used to simplify the object segmentation problem and to integrate the multiple views of the object into a single 3D representation. Symbolic labels are attached to objects once the temporal consistency of the sensed data is validated.

The planning level constructs actions plans to achieve pre-specified goals. The planner uses a high-level abstract symbolic model of the robot's environment based on an extension of the STRIPS language [95]. This model specifies the objects in the environment, their properties, and the actions that can be executed on those objects. Objects are represented simply by symbolic labels linked to the memory level OAC representations. Object and robot properties are specified by predicates and functions, such as $ingripper(x)$: the robot has grasped the object x in its gripper. Similarly, actions are represented in a high-level abstract manner as functions, such as $graspA-table(x)$ which corresponds to a grasp action directed at object x using grasp configuration A. It is the responsibility of the memory and sensorimotor levels to translate these symbolic action specifications into low-level motor control signals. The plans are constructed using PKS ("Planning with Knowledge and Sensing") [285, 286], a planner that can operate with incomplete information,

using a generalization of STRIPS [95]. Plans can be straightforward sequences of actions or they can include conditional branching based on the outcome of sensing actions. In this way, the planning level incorporates a form of reasoning on possible outcome of actions. It executes a plan by feeding the action primitives to the lower levels. Feedback from the lower levels allow the planning level to update its model of the state of the environment. This feedback allows the architecture to replan in the event of unexpected outcomes or invoke a some remedial action such as acquiring higher resolution scene representations.

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

11

Vernon et al.

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

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

Springer.com

A Roadmap for Cognitive Development in Humanoid Robots