

# Chapter 5

## Social Development of Artificial Cognition

Tony Belpaeme, Samantha Adams, Joachim de Greeff,  
Alessandro di Nuovo, Anthony Morse and Angelo Cangelosi

**Abstract** Recent years have seen a growing interest in applying insights from developmental psychology to build artificial intelligence and robotic systems. This endeavour, called developmental robotics, not only is a novel method of creating artificially intelligent systems, but also offers a new perspective on the development of human cognition. While once cognition was thought to be the product of the embodied brain, we now know that natural and artificial cognition results from the interplay between an adaptive brain, a growing body, the physical environment and a responsive social environment. This chapter gives three examples of how humanoid robots are used to unveil aspects of development, and how we can use development and learning to build better robots. We focus on the domains of word-meaning acquisition, abstract concept acquisition and number acquisition, and show that cognition needs embodiment and a social environment to develop. In addition, we argue that Spiking Neural Networks offer great potential for the implementation of artificial cognition on robots.

**Keywords** Human-Robot interaction · Social robots · Cognitive systems · Spiking artificial neural networks

### 5.1 Introduction

The recent, but fast expanding technological and financial investments in the production of intelligent robots rely on the design of robots with *effective and believable sensorimotor, cognitive and social capabilities*. For example, robots acting as assistive and social companions for the elderly must be able to autonomously navigate in the private home (or care home) where the elderly person lives, have fine manipulation skills to handle objects, be capable of understanding and using natural language for communication, and have believable social skills to enrich the experience of

---

T. Belpaeme (✉) · S. Adams · J. de Greeff · A. di Nuovo · A. Morse · A. Cangelosi  
Plymouth University, Centre for Robotics and Neural Systems, Plymouth, UK  
e-mail: tony.belpaeme@plymouth.ac.uk

© Springer International Publishing Switzerland 2016  
A. Esposito and L.C. Jain (eds.), *Toward Robotic Socially Believable Behaving Systems - Volume I*, Intelligent Systems Reference Library 105,  
DOI 10.1007/978-3-319-31056-5\_5

its elderly owner. Moreover, robots must be able to adapt to the requirements of the specific user, to react dynamically to changing environments and to learn new behavioural and cognitive skills through social interaction with the human user.

*Cognitive robotics* offers a feasible methodology for the design of robots with adaptive and learning capabilities, which can develop new skills through social interaction and learning. Cognitive robotics is a subfield of robotics in which robots are built based on insights gleaned from psychology, physiology and neuroscience, with the goal of replicating human-like performance on artificial systems [20, 78]. Cognitive robots are—as opposed to industrial robots—intended to work in open, unstructured and dynamic environments, the environments in which people typically feel at home, but robots do not. If someone asks a child to give a cup of water, the child can recognise and grasp the intended cup from among other objects, offer it, and do that while having a conversation. All this seems effortless to the child, but robots are—at this time—not able to do this in an open and dynamic environment. Robots might be programmed or trained to hand over a cup in a carefully controlled environment, but this would not generalise to handing over, say, a towel. As a rule of thumb, anything that seems effortless to humans is currently very hard for robots. And vice versa, we can build artificially intelligent systems and robots that can do things—such as playing chess or welding at a precise rate—that only very few of us ever master.

So why do classical approaches to building artificial intelligence and robots, that serve well to build chess playing computers and plan assembly tasks, fail to build AI that deals with unstructured and dynamic problems? The answer might lie in the study of development: young children seemingly effortlessly pick up skills which are very hard or impossible for robots to master. The question presents itself: can the same processes that are so successful in growing children be used to build intelligent robots? *Developmental robotics* is the interdisciplinary approach to the autonomous design of a complex repertoire of sensorimotor and mental capabilities in robots that takes direct inspiration from the developmental principles and mechanisms observed in the natural cognitive systems of children [7, 16, 79]. Developmental robotics relies on a highly interdisciplinary effort of empirical developmental sciences such as developmental psychology and neuroscience, and computational and engineering disciplines such as robotics and artificial intelligence. Developmental sciences provide the empirical bases and data to identify the general developmental and learning principles, mechanisms, models, and phenomena guiding the incremental acquisition of cognitive skills. The implementation of these principles and mechanisms into a robots control architecture and the testing through experiments where the robot interacts with its physical and social environment simultaneously permits the validation of such principles and the actual design of an increasingly complex set of complex behavioural and mental capabilities in robots.

Developmental robotics follows a series of general principles that characterise its approach to the design of intelligent behaviour in robots. Two of the key principles are the exploitation of embodiment factors in the development of cognitive capabilities and the focus on social learning.

Embodiment concerns the fundamental role of the body in cognition and intelligence. As Pfeifer and Scheier [57] claim, “intelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body” (p. 694). In psychology and cognitive science, the field of embodied cognition (also known as grounded cognition [10]) demonstrates the important roles of action, perception, and emotions in the grounding of cognitive functions such as memory and language [55]. For example, sensorimotor strategies, as postural changes, support the child in the early acquisition of words [63]. Gestures like pointing and finger counting are crucial in the acquisition of number knowledge [3]. Such developmental psychology studies are consistent with neuroscience evidence on embodied cognition, as brain-imaging studies showing that higher-order functions such as language share neural substrates normally associated with action processing [59]. The principle of social learning in developmental psychology is based on child development research on the role of social learning capabilities (instincts) in the very first days of life. This is evidenced for example by observations that newborn babies have an instinct to imitate the behavior of others and can imitate complex facial expressions after just few hours from birth [51]. Moreover, comparative psychology studies have shown that 18–24-month-old children have a drive to cooperate altruistically, a capacity missing in our closest genetic relatives as chimpanzees [80].

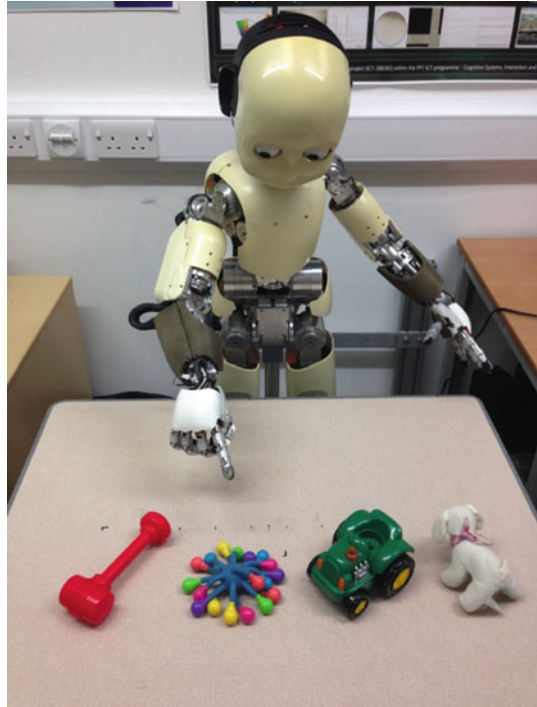
This chapter offers a summary of two recent studies on the modelling of embodiment and social learning in developmental humanoid robots. Both use the iCub humanoid platform both to exploit the properties of humanoid body configuration for embodiment modelling purposes and also for the benefits of using such humanoid platforms in social robotics scenarios. We will also discuss the potential of neuromorphic methods and hardware, as a first step for a brain-inspired approach to modelling the embodied basis of cognitive and communicative skills.

## 5.2 Why Embodiment Matters

Embodiment matters, not only in the development of natural cognition, but also in constructing *artificial* cognition. The brain or, in the case of robots, the control software cannot be seen as separate from the body in which it operates. Human cognition is deeply rooted in the shape of our bodies and how our bodies interact with the world. Likewise, when building artificial cognition, it is important to consider the full package of both the artificial intelligence inside the body interacting with the physical and social environment [4, 56].

In this chapter we provide an illustration of how an embodied perspective is used to imbue a robot with human-like skills. This requires a robot: we use the iCub platform, a child-sized humanoid robot specifically designed and built to facilitate developmental robotics [52] (see Fig. 5.1). In addition, an artificial cognitive model is required, which forms the theoretical backbone of the model.

**Fig. 5.1** The iCub robot learning to map words to objects. In this experiment, iCub knows linguistic labels for three of the four objects. In response to the question where is the dax (dax is a novel word), it points at the unknown object. With this iCub demonstrates “fast-mapping”, which is also observed in young children when they learn to map words to objects relying on only a few exposures and certain learning constraints [77]



### *5.2.1 The Origins of Abstract Concepts and Number: A Detailed Study*

Recent studies have proposed that multiple representational systems, involving both sensorimotor as well as linguistic systems, might be playing a primary role in how children acquire abstract concepts and words (e.g. [48]). Theories such as the LASS theory [11], according to which both the linguistic system as well as the sensorimotor system (through simulation) are activated in the processing of word meaning to different degrees under different task conditions, and the WAT (Words as Tools) approach proposed by Borghi and Cimatti [13], have suggested and furnished evidence on the synergetic role both language and sensorimotor experience play in the acquisition of abstract concepts, and on how important the modality by which these words are learned is.

Finger counting has been shown to have an important role in the development of number cognition [3, 30]. Embodied cognition researchers find this innate ability particularly interesting, because of the sensorimotor contribution that it makes to the development of numerical cognition, and some consider it as “the most prominent example of embodied numerical cognition” [9]. Evidence coming from developmental, neurocognitive as well as neuroimaging studies suggest that finger counting activity helps build motor-based representations of number that continue to influence

number processing well into adulthood, indicating that abstract cognition is rooted in bodily experiences [33]. These motor-based representations have been argued to facilitate the emergence of number concepts through a bottom-up process, starting from sensorimotor experiences [5].

In our view, finger counting, can also be seen as a means by which direct sensory experience can serve the purpose of grounding number words as well as numerical symbols, initially as low level symbols from the combination of already grounded ones, something known as grounding transfer [15, 40].

A number of connectionist models have simulated different aspects of number learning. A multi neural net approach was presented in [2] to explore quantification abilities and how they might arise in development, using a combination of supervised and unsupervised networks and learning techniques to simulate subitization (the phenomenon by which subjects appear to produce immediate quantification judgements, usually involving up to four objects, without the need to count them) and counting. The authors used a combined and modular approach, providing a simulation of different cognitive abilities that might be involved in the cognition of number (each of which would have their own evolutionary history in the brain), and is in keeping with Dehaenes triple code model [25]. In [60], using a hybrid artificial vision connectionist architecture, authors targeted aspects of language related to number such as linguistic quantifiers. They ground linguistic quantifiers such as few, several, many, in perception, taking into consideration contextual factors. Their model, after being trained and evaluated with experimental data using a dual-route neural network, is able to count objects in visual scenes and select the quantifier that best describes the scene.

Not many robotics studies have attempted to extend this. A cognitive robotics paradigm was used in [61, 62], where the authors explored embodied aspects of mathematical cognition such as the interactions between numbers and space, reproducing three psychological phenomena connected with number processing, namely size and distance effects, the SNARC effect and the Posner-SNARC effect.<sup>1</sup> The focus was on counting and on the contribution of counting gestures such as pointing. These models, however, did not consider the role of finger counting in numerical abilities.

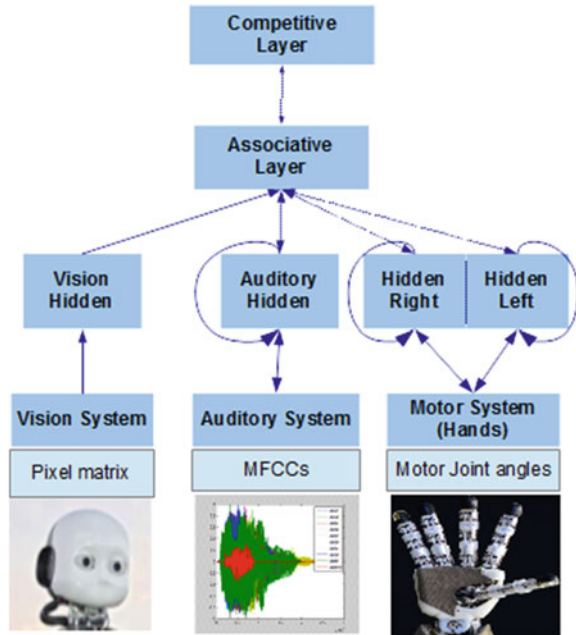
Using a cognitive developmental robotics paradigm we explore whether finger counting and the association of number words (or tags) to the fingers, could serve to bootstrap the representation of number in a cognitive robot [23, 31, 32]. Our embodied robot experiments indicate that aspects of the development of this knowledge can be accounted for not only by way of bodily representations, but that a relatively simple artificial neural network is sufficient to achieve this.

The complete architecture proposed is shown in Fig. 5.2: the lower layer contains the motor controller/memory, and the auditory and the vision sub-systems. These are directly connected to the robotic platform. In the upper part there are the units

---

<sup>1</sup>SNARC, spatial-numerical association of response codes, is the effect whereby quantities seem to be spatially organised. People respond faster to small numbers with their left hand, and respond faster to large numbers with their right hand.

**Fig. 5.2** Schematic of the robots cognitive system for number cognition



with abstract functions: the associative network and the competitive layer classifier. Note that the recurrent system's external inputs coincide with the outputs, indeed proprioceptive information from the motor and auditory systems is an input for the system during the training phase, while it is the control output when the system is operating.

Inputs are the joint angles, read from the encoders of the iCub hands, the mel-frequency cepstral coefficients (MFCC) to represent each number word from one to ten, and digits of  $5 \times 2$  black and white pixels to represent number symbols. All numbers are in the range  $[-1, 1]$ . For number symbols, each element can be either  $-1$ , when the pixel is white, or  $+1$  when the pixel is black.

The role of the competitive layer classifier is to simulate the final processing of the numbers, after a number is correctly classified into its class, the appropriate action can be started, e.g. the production of the corresponding word, of a symbol, the manipulation of an object and so on. The competitive layer classifier is implemented using the softmax transfer function that gives as output the probability/likelihood of each classification. We ensure that all of the output values are between 0 and 1, and that their sum is 1. The Switch/Associative Layer operates as a feedback system with the possibility to start and reset the motor/auditory layers and to derive the activations of one layer from the ones of the other.

Several experiments are run with the above architecture using the iCub robotic platform. In the first experiment [23], the main goal is to test the ability of the proposed cognitive system to learn numbers by comparing the performance of different

ways of training the number knowledge of the robot with: (1) the internal representation (hidden units activation) of a given finger sequence; (2) the MFCC coefficients of number words out of sequence; (3) the internal representation of the number words sequence; (4) the internal representation of finger sequences plus the MFCC of number words out of sequence (i.e. learning words while counting); (5) internal representations of the sequences of both fingers and number words together (i.e. learning to count with fingers and words).

Looking at the developmental results, we again see that number words learnt out of sequence are the least efficient to learn. Conversely, if number words are learnt in sequence and internal representations are used as inputs, the learning is faster in terms of precision of classification, but is not as strong as when the learning involves also the use of fingers. Indeed, best results are obtained when internal representation of words and fingers are used together as input (Figs. 5.3 and 5.4).

A second experiment [31] focuses on learning associations between the internal representations (i.e. hidden unit activations) of number digits and the number words. Abstract concepts like the written representation of numbers is an important milestone in the child's unfolding cognitive development [81]. The young math learner must make the transition from a concrete number situation, in which the counting of objects (with fingers often being the first), to that of using a written form to stand for the quantities the sets of objects come to represent. This already challenging process

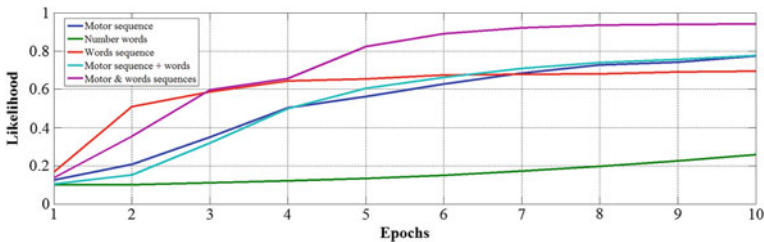


Fig. 5.3 Average likelihood with number classes with varying epochs

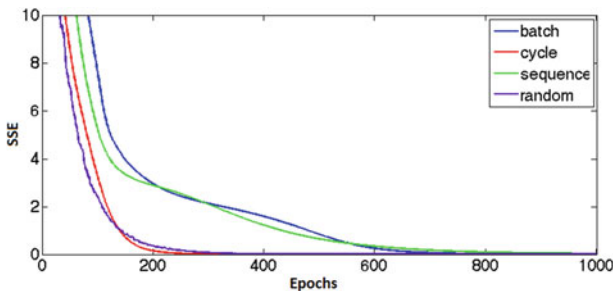


Fig. 5.4 Developmental learning of the association between number words and digits. Four different weight training methods are compared

is often coupled to that of learning a verbal number system, which depending on the particular language being used is not always transparent to children.

In this experiment four training strategies are considered: **Batch**, network weights are updated at the end of an entire pass through the input data; **Incremental** (3 strategies), network weights are learned with incremental updates after each presentation of an input order. Inputs are presented in **sequential** (i.e. from 1 to 10 each epoch), **random** (the order is randomly shuffled at each epoch), or **cyclic order** (the order is shifted after each epoch). Hidden unit activations are evaluated from the network with the best (lowest value of) performance function.

From this study we can conclude that the batch learning and the sequential strategies are less effective compared to others. They are slower to learn (i.e. they need more epochs) the final error (measured as sum of squared errors, or SSE) is several orders of magnitude higher than for random and cyclic order incremental training.

Once the number sequences are learnt, an interesting feature of the proposed cognitive system is the possibility to easily build up the ability to manipulate numbers with the development of the switch-associative network. Indeed, this ability can be modelled by extending the capabilities of the associative network from the simple start and stop, to its transferring and mapping to the basic operation of addition. The operation of addition can be seen as a direct development of the concurrent learning of the two recurrent units (motor and auditory). Indeed, if one of the two does the actual counting of the operands, the other can be used as a buffer memory to add the result, when it is done, the final number can be transferred from the buffer to the other unit and then inputted to the final processor (the classifier in our system). Here we want to build on this to show how the proposed architecture can take advantage of the previously learnt capability.

As an example let us consider  $2+2$ , in this case the following steps will be taken:

1. The first operand is recognised by the visual system and, thanks to the associative network, the auditory internal representation is activated.
2. Auditory and motor networks will count until the corresponding activation of number 2 is reached. This step corresponds to the idle, start, counting (cycled twice) then done statuses of the associative network.
3. The sum operator is recognised so the associative layer resets the auditory network, while the first operand remains stored in the motor memory.
4. The second operand is recognised by the visual system, so the other networks restart counting as in step 1, until the auditory network reaches the activation corresponding to the number 2. In the meantime, the motor network reaches the activation of the number 4.
5. After the auditory network stops, the associative network recognises that the work is done so the total (4) is incepted from the fingers network to the auditory network thanks to the associative connection.
6. Finally the output of the resulting number (4) is produced for final processing (in our case the classifier).

The steps are depicted in Fig. 5.5.



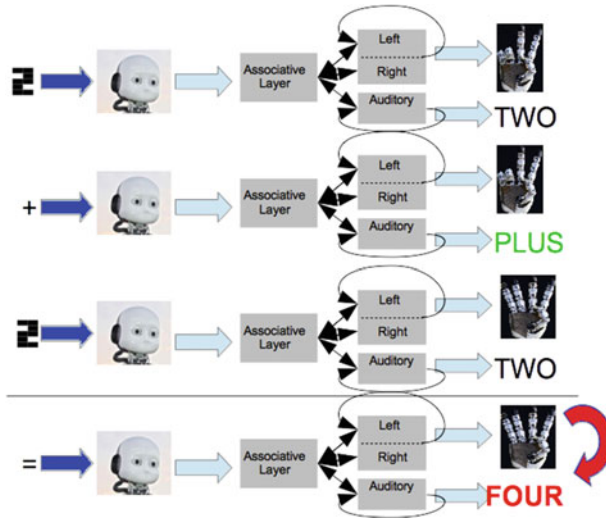


Fig. 5.5 Example of the execution of the operation of sum

### 5.3 Learning Through Social Interaction

As argued in the previous section, the seat of cognition is not the brain, but instead cognition emerges from the interaction between the brain, the body and the physical environment. While this holds for cognitive development of most animals, this picture is incomplete for some social species, and most significantly it is incomplete for humans. For human cognition to develop, one last element is required, namely the *social environment*. When including the social environment in cognition, this is sometimes known as “extended cognition”<sup>2</sup> [64].

While some elements of human cognition in all likelihood develop without input from the social environment (grasping and manipulation, for example, most likely develop without relying on social interaction), human infants grow up in a rich social environment. In this environment, social input in various shape and form is offered to the child, impacting on its cognitive development. Children learn from observing others: mimicry and imitation are potent mechanisms for acquiring cognitive skills [54] and rely on more skilled others to ostensibly demonstrate a skill, which is then imitated by the learner. Quite often the demonstration will be tailored as to promote successful interpretation of the demonstration by the young learner; demonstrating more slowly or emphasising salient elements of the skill to be acquired. The demonstrator also is able to provide feedback on the success of the demonstration, and can actively correct elements of the skill that are not yet fully established. Imitation, in

<sup>2</sup>Not to be confused with the Extended Mind hypothesis, in which cognition is argued to extend to the external world. As such external objects, such as canes, notepads and calculators, are seen as being integral to human cognition [21].

some form or other, is observed in many animals—primates and birds are known to imitate extensively—and as such imitation is a form of social learning that is not uniquely human [41]. However, *language is uniquely human*. While many species communicate, no other species has access to the open communication system that language is.

It has been claimed that language is such a hard problem that is unlearnable, and can only result from an innate language of thought pre-specified by genetic evolution [19, 35, 36, 58]. How else could the child, a passive observer viewing a cluttered scene, know to which feature or collection of features a spoken word referred? By contrast, embodiment views the learning child as anything but passive [73]; their attention is clearly focused and they are ‘doing’ (reaching, holding, banging, manipulating...) sometimes being physically lead by the caregiver [82]. Words are not merely spoken either, child directed speech is not simply speech directed at the child, but is manipulated to highlight events, and is just as much directed by the child’s attention and reaction. Smith et al. [68] go further, highlighting just how dominant a held object is in the infants’ field of view. From this perspective the language learning child is not really aware of all the perceptual clutter (the held object is simply occluding most of it), and the spoken words often relate to what the child is currently doing, holding or attending to. As such, the learning of simple concrete item-based word-object and word-action mappings becomes possible, and we have demonstrated the basic principles involved on robotic systems [53, 77].

Moving beyond simple word-object mappings, Tomasello [75] further highlights that from a concrete item-based vocabulary children gradually (over many years) develop the ability to construct more abstract and adult-like linguistic constructions. This gradually increasing complexity of language presents a significant challenge to the hypothesis that language is innate. We therefore suggest that language is not symbol manipulation in the head but is a sensorimotor process, whereby words prime or predict associated features (be they combinations of sensory features, motor actions, or affordances), and likewise these features prime their associated words.

Language has many functions, next to its obvious function as communication system, it also supports cognition in ways that are not always recognised. One is that language is used in concept acquisition. Humans are, in the words of Terrence Deacon, a “symbolic species” [24]. We cut up our perceptual experience in concepts, and can subsequently order these concepts into hierarchical concepts. Concepts allow us to reason and are the brain’s way of compressing sensory input into fewer, finite units. When we assign a linguistic label to a concept (a word, utterance or linguistic construction), that concept can be communicated to others. Concepts are central to human cognition, but it is not clear where concepts come from. Are they learnt by a child when growing up? Are some concepts innate, and some learnt? If so, which concepts are innate and which are learnt? And, when concepts are acquired, how exactly are they acquired?

This latter question is important: *how can a child acquire concepts?* And by extension, can similar processes be used to let an artificial system, such as a robot, acquire human-like concepts? Sect. 5.2.1 shows how linguistic labels (i.e. words) can

be used to bind external perception with internal representations. In this section we look into the contribution of the social interaction on word-meaning acquisition.

As children develop, there is a rich and frequent interaction between the child and the environment. Not only the physical environment is explored, but also the social environment. And while the physical environment does not actively respond to actions by the child, the social environment (i.e. the child's siblings, parents or other carers) do actively respond to the developing child. In language learning, phenomena such as infant-directed speech—where carers address the child in a simpler and hyperarticulated language—aids language acquisition. Likewise, when learning the semantics of language, the carer-child dyad often engages in rich interactions in which joint attention and deictic pointing is combined with the naming of objects or actions. Together with a number of learning biases [76], this enables the child to rapidly acquire a set of words and semantic associations [50].

Inspired by these observations, we explore if similar mechanisms could be used to accelerate robot learning. In our study, the robot learns from people in a way that is similar to how children learn from others. In this *socially guided machine learning* [47] the machine is not offered a batch of training data to learn from, but instead engages in a high-resolution interaction which a human, whereby the machine invites the human to offer tailored training data to optimise its learning.

We focus on the task of learning associations between words and referents [12], for which the learner has to construct internal representations linking linguistic symbols with external referents. These dynamics of meaning acquisition have been explored in detail, often using simulations in which agents bootstrap a shared symbol system and meanings—e.g. [70–72]. However, in these simulations the agents do not actively influence the learning process by querying their social environment. Early simulations have shown that active learning can result in improved performance [29]. When an agent uses strategies to actively elicit better training examples from other agents, the learner learns faster and better. The strategies consist of active learning (whereby the learner points out a referent in the world which it would like to know the linguistic label for, similar to a child pointing out something in the presence of a carer), knowledge querying (whereby the learner verifies its internal knowledge by using it and asking the carer for feedback, mimicking the way in which children name objects in their environment and invite adults to correct them or confirm their linguistic labels) and contrastive learning (in which an association between a word and a referent is increased, but association between that word and other referents is decreased).

While the strategies result in better learning in simulation, we wish to confirm if this would still hold in the real world: where a social robot is learning from a human tutor. To this end we design a setup in which a social robot sits across a human tutor (see Fig. 5.6). Between the robot and human, a touchscreen is placed through which the interaction takes place. The participants are asked to teach the robot the concepts of animal classes (mammal, insect, invertebrate, ...), using images of animals (e.g. a bear, an ant, a lizard, ...).

The robot uses learning strategies identical to the simulation model [29], the contribution of the social robot setup is on the one hand the learning from people



**Fig. 5.6** Experimental setup, with the social robot sitting across the participant. The participant is invited to teach the robot to correctly match images of animals with animal classes

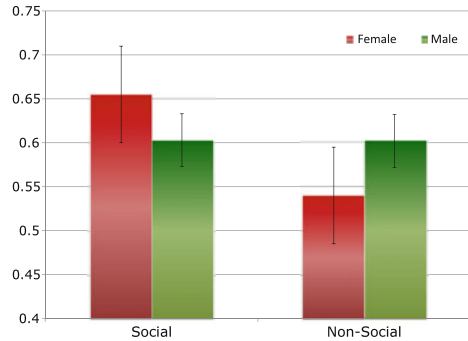
rather than from other simulated agents, and on the other hand the introduction of additional communication channels, such as eye gaze and affective communication. To aid social communication and to invite people to help the robot learn, the robot is deliberately designed to resemble a young child [26].

The experiment uses two conditions: in one condition people interact with a social robot, using the learning strategies detailed above and using congruent linguistic and facial expressions to support the active learning, in the second condition, the robot learns, but does not use any of the above strategies to learn more efficiently; we refer to this condition as the “non-social robot”. 19 subjects interacted with the social robot, and 20 with the non-social robot condition. Full details can be found in [28].

Results show that in both conditions, the robot learns to correctly match instances of animals to animal classes, illustrating that the learning algorithms works as expected. The social robot learning is faster and slightly better than that of the non-social robot, as predicted by the simulation results. It is interesting to observe that there is a marked gender effect in the results: female participants achieve a significantly higher learning success when interacting with the social robot, and this drops significantly for the non-social robot. Male participants achieve similar learning results for both conditions (see Fig. 5.7). This suggests that female participants in our study are more sensitive to the social cues expressed by the robot, while this is not the case for the male participants.

Finally, a careful analysis of the data shows that people readily form a “mental model” of the robot: both in the social and non-social condition people will offer training data to the robot that are tailored to the current performance of the robot [28], thereby showing that people form a model of the robot’s mental state.

**Fig. 5.7** Word-meaning learning of the social and non-social robot; note how the social robot learns more from female participants in our experiment, while it learns significantly less from male participants. Error bars are 95 % confidence intervals



This experiment convincingly illustrates that social robots can elicit better training experiences. The careful design of the appearance and the behaviour of the robot can lead to improved learning on robots, and taps into the human propensity for tutoring.

## 5.4 Powering Artificial Cognition with Spiking Neural Networks

The desire to endow robots with sophisticated human-like capabilities raises some major challenges as traditional computing and engineering approaches can only achieve so much. They can and have been used to mimic human capabilities to various levels of abstraction but it is difficult to make artificial systems that behave in the same way as natural ones do if we do not fully understand all the neural processing which generates our own behaviour. It is also often difficult to translate biological concepts into a traditional computing/engineering framework without making severe compromises. The sensory pre-processing and higher level cognitive processing that is required to achieve such human-like learning capabilities in an embodied developmental robotics scenario likely requires significant computing power which is in conflict with the limited energy resources available on an autonomous robot. It should be noted, however, that natural neural systems manage to operate in real time, be fault tolerant and flexible despite having very low power requirements. Therefore, it seems logical to explore more in depth bio-inspired approaches to robotics. For example, where artificial brains and nervous systems are implemented using techniques inspired by greater understanding of how real neurons work. Arbib et al. [6] defines the field of Neurorobotics as

... the design of computational structures for robots inspired by the study of the nervous system of humans and other animals.

and suggests that neural models more closely matching the biology may more clearly reveal the computational principles necessary for cognitive robotics while illuminating human (and animal) brain function.

In parallel, the field of Computational Neuroscience (the study of brain function using biologically realistic models of neurons over multiple scales from single neuron dynamics up to networks of neurons) has made considerable progress on spiking neuron based models of sensory and cognitive processes in the mammalian neo-cortex. Spiking Neural Networks (SNNs) are the “third generation” of Neural Networks [49] and mimic how real neurons compute: with discrete pulses rather than a continuously varying activation. The spiking neuron is, of course, still an abstraction from a real neuron but depending upon the application and required level of biological detail, there are various types of spiking neuron model to choose from. However, there is also a trade-off between the level of biological detail and computational overhead (for a review and discussion see [42]).

In neurobiological experimental studies neuron responses have been predominantly measured as a spike rate, however there is accumulating evidence that spike timing is also important. Experimental evidence exists for fast processing (occurring within 100 ms of an image presentation) in the human visual system [74] which implies that spike timing information may be more important than spike rates as there is not enough time to generate a meaningful spike rate in very short time intervals. Spike timing also seems to be important in learning: Spike Timing Dependent Plasticity (STDP) is a currently favoured model for learning in real neurons. Experimental and modelling studies have shown that this form of Hebbian plasticity, where the relative firing times of pre and post-synaptic neurons influence the strengthening or weakening of connections, is the mechanism that real neurons use [69]. When firing times are causally related (i.e. the pre-synaptic spike is emitted before the post-synaptic spike) then the synapse is strengthened (Long Term Potentiation or LTP). When firing times are not causally related (i.e. the post-synaptic spike occurs before the pre-synaptic one) then the synapse is weakened (Long Term Depression or LTD).

Of particular relevance to modelling human cognitive function, some neurobiological experiments have suggested that spike-timing is also directly important at the cognitive/behavioural level as well as in learning [8, 66].

There have been a few research projects involving robotic implementations based upon human-like capabilities using spiking neural networks. Three notable examples are the Darwin series of robots [34, 44], the humanoid CRONOS/SIMNOS project [38] and the control of an iCub arm with an SNN and STDP [14].

The iSpike API [39] has a lot of promise for facilitating interfacing between SNNs and humanoid robots but as yet no practical demonstrations exist. Therefore, it is only relatively recently that works using spiking neural networks for practical humanoid robotics applications have begun to emerge. Certainly advances in software and hardware over the last ten years or so have made SNNs an increasingly feasible option for robotics applications. On the software side several general purpose spiking neuron simulators are freely available which means that researchers do not have to code a modelling framework from scratch, and they also benefit from a community of users using the same tool. Desktop computing hardware is now available that can perform parallel processing (e.g. GPU) at an affordable price. But this can only take us so far. Until now, in practice most Neurorobotic systems, e.g. Chersi [18] have

simulated the neural component on an external host PC which limits the ability of the robot to truly perform autonomously in real time.

The emerging field of Neuromorphic Engineering is making it possible to simulate large neural networks in hardware in real time. Neural chips are massively parallel arrays of processors that can simulate thousands of neurons simultaneously in a fast, energy efficient way, thus making it possible to move neural applications on board robots. This technology is currently being employed in dedicated hardware devices to perform specific bio-inspired functions, for example, the asynchronous temporal contrast silicon retina [27] and the silicon cochlea [17]. There have also been several larger-scale projects for general purpose brain modelling. For example, the CAVIAR project; a massively parallel hardware implementation of a spike-based sensing-processing-learning-actuating system inspired by the physiology of the nervous system [65], the FACETS project (completed in 2010) delivering both neuromorphic hardware and software and the NeuroGrid project at Stanford which has developed a hybrid analogue-digital neuromorphic solution capable of modelling up to 1 million neurons (reviewed in [67]). More recently, the SpiNNaker project has delivered a state-of-the-art real-time neuromorphic modelling environment that can be scaled-up to model up to a billion point-neuron models [43].

The parallel advances in computational neuroscience and in the hardware implementation of large-scale neural networks, provide the opportunity for an accelerated understanding of brain functions and for the design of interactive robotic systems based on brain-inspired control systems. However, currently there are very few practical robotics implementations using neuromorphic systems. Two notable works are [46] which developed a solution using both a silicon retina, an FPGA and neuromorphic hardware to enable a humanoid robot to point in the direction of a moving object, and, more recently [22] which developed a line following robot using a silicon retina and a prototype 4-chip SpiNNaker neuromorphic board.

Adams et al. [1] recently introduced a Neurorobotics system integrating the humanoid iCub robot and a SpiNNaker neuromorphic board to solve a behaviourally relevant task: goal-directed attentional selection. Using an enhanced version of an existing SNN model with layers inspired by real brain areas in the mammalian visual system [37], iCub was equipped with the ability to fixate attention upon a selected stimulus. Although in this particular implementation the selected or “preferred” stimulus was fixed in advance the network has the option to enable STDP learning to learn the preferred stimulus.

This study demonstrated the first steps in creating a cognitive system incorporating several important features for prospective Neurorobots:

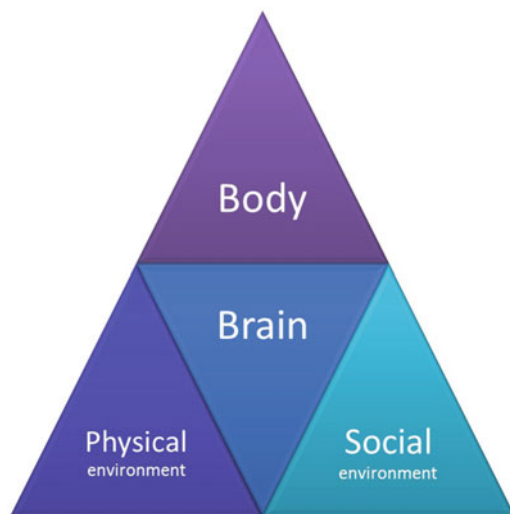
1. Universally configurable hardware that can run a variety of SNNs.
2. Standard interfacing methods that eliminate difficult low-level issues of connectors, cabling, signal voltages, and protocols.
3. Scalability—the SpiNNaker hardware is designed to be able to run very large SNNs and the optimal placement of networks onto the hardware is abstracted away from the user.

More work needs to be done to develop practical applications that have a solid biologically-inspired theoretical basis and which can be scaled up and transferred seamlessly to run on neuromorphic hardware to take advantage of their specialist processing capabilities and low power requirements. For realistically large and effective SNNs to become possible in robotic hardware, ensuring that future neural models and simulations are actually implementable in neuromorphic hardware is important. It is also important to develop models which challenge the capabilities of such hardware and stimulate further developments.

## 5.5 Conclusion and Outlook

The studies described here illustrate how artificial cognition, just as its natural counterpart, benefits from being grounded and embodied. This occurs at several levels: the body of the cognitive system shapes its cognition, but so does the physical environment and the social environment. Human-like cognition results from the tight interaction between these four constituents, see Fig. 5.8. When one of the four constituents is missing, it is still possible to recreate certain aspects of cognition. For example, the social component is missing in much of animal cognition and some aspects of human cognition—such as manipulation or locomotion—can develop in the absence of the social environment. Or when the body is missing, systems have been shown to still be able to reach human levels of performance on specific tasks. Latent Semantic Analysis, for example, is able to pass synonymy tests just using statistical co-occurrence information of words in large text corpora [45]. However, to replicate natural human cognition in its full scope, we make argue that all four components—body, brain,

**Fig. 5.8** The four constituents of human cognition: the brain, the body, and the physical and social environment. We argue that no human-like cognition can develop without the presence of these four components





physical environment and social environment—are required, and that all four cannot be seen as separate entities, rather they intertwine and operate in close association with each other. In addition, we believe that a thorough understanding of the neural processes underpinning natural cognition will aid in the design and implementation of artificial equivalents; key to this might be spiking neural networks.

## References

1. Adams S, Rast A, Patterson C, Galluppi F, Brohan K, Perez-Carrasco JA, Wennekers T, Furber S, Cangelosi A (2014) Towards real-world neurobotics: Integrated neuromorphic visual attention. In: Proceedings of 21st international conference on neural information processing (ICONIP), pp 563–570
2. Ahmad K, Casey M, Bale T (2002) Connectionist simulation of quantification skills. *Connect Sci* 14(3):165–201
3. Alibali MW, DiRusso AA (1999) The function of gesture in learning to count: more than keeping track. *Cognit Dev* 14(1):37–56
4. Anderson ML (2003) Embodied cognition: a field guide. *Artif Intell* 149(1):91–130
5. Andres M, Di Luca S, Pesenti M (2008) Finger counting: the missing tool? *Behav Brain Sci* 31(06):642–643
6. Arbib MA, Metta G, van der Smagt P (2008) Neurobotics: From vision to action. In: Khatib O Siciliano B (eds) Springer Handbook of Robotics, Springer-Verlag, pp 1453–1480
7. Asada M, Hosoda K, Kuniyoshi Y, Ishiguro H, Inui T, Oginio Y, Yoshida C (2009) Cognitive developmental robotics: a survey. *IEEE Trans Auton Mental Dev* 1(1):12–34
8. Ayzenshtat I, Meirovithz E, Edelman H, Werner-Reiss U, Bienenstock E, Abeles M, Slovin H (2010) Precise spatiotemporal patterns among visual cortical areas and their relation to visual stimulus processing. *J Neurosci* 40:11232–11245
9. Bahnmüller J, Dresler T, Ehrlis AC, Cress U, Nuerk HC (2014) Nirs in motion unraveling the neurocognitive underpinnings of embodied numerical cognition. *Front Psychol* 5:743
10. Barsalou LW (2008) Grounded cognition. *Annu Rev Psychol* 59:617–645
11. Barsalou LW, Santos A, Simmons WK, Wilson CD (2008) Language and simulation in conceptual processing. Symbols, embodiment, and meaning pp 245–283
12. Bloom P (2000) How children learn the meanings of words. The MIT Press, Cambridge
13. Borghi AM, Cimatti F (2012) Words are not just words: the social acquisition of abstract words. *Rivista Italiana di Filosofia del Linguaggio* 5:22–37
14. Bouganis A, Shanahan M (2010) Training a spiking neural network to control a 4-dof robotic arm based on spike timing-dependent plasticity. *Proc IJCNN* 2010:1–8
15. Cangelosi A, Riga T (2006) An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. *Cognit Sci* 30(4):673–689
16. Cangelosi A, Schlesinger M (2015) Developmental robotics: from babies to robots. The MIT Press, Cambridge
17. Chan V, Liu SC, van Shaik A (2007) Aer ear: a matched silicon cochlea pair with address event representation interface. *IEEE Trans Circuits Syst I: Spec Issue Smart Sens* 54:48–49
18. Chersi F (2012) Learning through imitation: a biological approach to robotics. *IEEE Trans Auton Mental Dev* 4(3):204–214
19. Chomsky N (1995) The minimalist program. Cambridge Univ Press, Cambridge
20. Christaller T (1999) Cognitive robotics: a new approach to artificial intelligence. *Artif Life Robot* 3(4):221–224
21. Clark A, Chalmers D (1998) The extended mind. analysis pp 7–19
22. Davies S, Patterson C, Galuppi F, Rast A, Lester D, Furber S (2010) Interfacing real-time spiking i/o with the spinnaker neuromimetic architecture. In: Proceedings 17th international

- conference, ICONIP 2010: 17th international conference, ICONIP 2010: Australian journal of intelligent information processing systems, vol 11, pp 7–11
23. De La Cruz VM, Di Nuovo A, Di Nuovo S, Cangelosi A (2014) Making fingers and words count in a cognitive robot. *Front Behav Neurosci* 8:1–12
  24. Deacon TW (1997) *The symbolic species: the co-evolution of language and the brain*. Norton, New York
  25. Dehaene S (2000) *The cognitive neuroscience of numeracy: exploring the cerebral substrate, the development, and the pathologies of number sense*. Scientific research faces a new millennium, Carving our destiny
  26. Delaunay F, de Greeff J, Belpaeme T (2010) A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. *Proceedings of the 5th ACM/IEEE international conference on human-robot interaction (HRI2010)*, Mar 2–5 (2010). IEEE Press, Osaka, Japan, pp 39–44
  27. Delbruck T (2008) Frame-free dynamic digital vision. In: *Proceedings of international advanced electronics for quality life and society, symposium on secure-life electronics*, pp 21–26
  28. de Greeff J, Belpaeme (2015) Why robots should be social: Enhancing machine learning through social human-robot interaction. *PLOS One* In press
  29. de Greeff J, Delaunay F, Belpaeme T (2009) Human-robot interaction in concept acquisition: a computational model. In: Triesch J, Zhang Z (eds) *IEEE international conference on development and learning (ICDL 2009)*. IEEE, Shanghai
  30. Di Luca S, Pesenti M (2011) Finger numeral representations: more than just another symbolic code. *Front Psychol* 2:272
  31. Di Nuovo A, De La Cruz VM, Cangelosi A (2014a) Grounding fingers, words and numbers in a cognitive developmental robot. In: *IEEE symposium on computational intelligence, cognitive algorithms, mind, and brain (CCMB, 2014)*. IEEE, pp 9–15
  32. Di Nuovo A, De La Cruz VM, Cangelosi A, Di Nuovo S (2014b) The icub learns numbers: An embodied cognition study. In: *International joint conference on neural networks (IJCNN, 2014)*. IEEE, pp 692–699
  33. Domahs F, Moeller K, Huber S, Willmes K, Nuerk HC (2010) Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116(2):251–266
  34. Edelman G (2007) Learning in and from brain-based devices. *Science* 318:1103–1105
  35. Fodor JA (1975) *The language of thought*. Harvard University Press, Cambridge
  36. Fodor JA (2008) *LOT 2: the language of thought revisited: the language of thought revisited*. Oxford University Press, Oxford
  37. Galluppi F, Brohan K, Davidson S, Serrano-Gottarredona T, Corasco JAP, Linares-Barranco B, Furber S (2012) A real-time, event driven neuromorphic system for goal-directed attentional selection. In: *ICONIP 2012*
  38. Gamez D, Newcombe R, Holland O, Knight R (2006) Two simulation tools for biologically inspired virtual robotics. In: *Proceedings of the IEEE 5th chapter conference on advances in cybernetic systems*
  39. Gamez D, Fidjeland A, Lazdins E (2012) Iispik: a spiking neural interface for the icub robot. *Bioinspir Biomim* 7(2):025008
  40. Harnad S (1990) The symbol grounding problem. *Phys D: Nonlinear Phenom* 42(1):335–346
  41. Hurley SL, Chater N (2005) *Perspectives on Imitation: mechanisms of imitation and imitation in animals*, vol 1. MIT Press
  42. Izhikevich EM (2004) Which model to use for cortical spiking neurons? *IEEE Trans Neural Netw* 15(5):1063–1070
  43. Jin X, Lujan M, Plana L, Davies S, Temple S, Furber S (2010) Modeling spiking neural networks on spinnaker. *Comput Sci Eng* 21(5):91–97
  44. Krichmar J, Edelman G (2003) Brain-based devices: intelligent systems based on principles of the nervous system. In: *Proceedings of the 2003 IEEE/RSJ international conference on intelligent robots and systems vol 1*

45. Landauer TK, Dumais ST (1997) A solution to plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev* 104(2):211–240
46. Linares-Barranco A, Gomez-Rodriguez F, Jimenez-Fernandez A, Delbr-ck T, Lichtensteiner P (2007) Using fpga for visuo-motor control with a silicon retina and a humanoid robot. In: *Proceedings of the IEEE symposium on circuits and systems (ISCAS 2007)*, pp 1192–1195
47. Lockerd A, Breazeal C (2004) Tutelage and socially guided robot learning. In: *Proceedings of IEEE/RSJ international conference on intelligent robots and systems (IROS 2004)*
48. Louwerse MM, Jeuniaux P (2010) The linguistic and embodied nature of conceptual processing. *Cognition* 114(1):96–104
49. Maass W (1997) Networks of spiking neurons: the third generation of neural network models. *Neural Netw* 10:1659–1671
50. Markman EM (1989) Categorization and naming in children: problems of induction. The MIT Press, Cambridge
51. Meltzoff AN, Moore MK (1994) Imitation, memory, and the representation of persons. *Infant Behav Dev* 17(1):83–99
52. Metta G, Sandini G, Vernon D, Natale L, Nori F (2008) The icub humanoid robot: an open platform for research in embodied cognition. In: *Proceedings of the 8th workshop on performance metrics for intelligent systems, ACM*, pp 50–56
53. Morse AF, Belpeame T, Cangelosi A, Floccia C, Carlson L, Hoelscher C, Shipley T (2011) Modeling u-shaped performance curves in ongoing development. In: *Expanding the space of cognitive science: proceedings of the 23rd annual meeting of the cognitive science society*
54. Nehaniv CL, Dautenhahn K (2002) Imitation in animals and artifacts. MIT Press, Cambridge
55. Pecher D, Zwaan RA (2005) *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press, Cambridge
56. Pfeifer R, Bongard J (2006) *How the body shapes the way we think: a new view of intelligence*. MIT press, Cambridge
57. Pfeifer R, Scheier C (1999) *Understanding intelligence*. The MIT Press, Cambridge
58. Pinker S (1994) *The language instinct: how the mind creates language*. W. Morrow, New York
59. Pulvermüller F (2005) Brain mechanisms linking language and action. *Nat Rev Neurosci* 6(7):576–582
60. Rajapakse RK, Cangelosi A, Coventry KR, Newstead S, Bacon A (2005) Connectionist modeling of linguistic quantifiers. In: *Artificial neural networks: formal models and their applications-ICANN 2005*, Springer, pp 679–684
61. Rucinski M, Cangelosi A, Belpaeme T (2011) An embodied developmental robotic model of interactions between numbers and space. *Expanding the space of cognitive science: proceedings of the 23rd annual meeting of the cognitive science society*. Cognitive Science Society Austin, TX, pp 237–242
62. Rucinski M, Cangelosi A, Belpaeme T (2012) Robotic model of the contribution of gesture to learning to count. In: *IEEE International conference on development and learning and epigenetic robotics (ICDL, 2012)*. IEEE, pp 1–6
63. Samuelson LK, Smith LB, Perry LK, Spencer JP (2011) Grounding word learning in space. *PLOS One* 6(12):e28,095
64. Seabra Lopes L, Belpaeme T, Cowley S (2008) Beyond the individual: new insights on language, cognition and robots. *Connect Sci* 20(4):231–237
65. Serrano-Gotarredona R, Oster M, Lichtsteiner P, Linares-Barranco A, Paz-Vicente R, Gomez-Rodriguez F, Camunas-Mesa L, Berner R, Rivas M, Delbr-ck T, Liu SC, Douglas R, Hafliger P, Jimenez-Moreno G, Civit A, Serrano-Gotarredona T, Acosta-Jimenez A, Linares-Barranco B (2009) Caviar: A 45k-neuron, 5m-synapse, 12g-connects/sec aer hardware sensory-processing-learning-actuating system for high speed visual object recognition and tracking. *IEEE Trans Neural Netw* 20:1417–1438
66. Shmiel T, Drori R, Shmiel O, Ben-Shaul Y, Nadasdy Z, Shemesh M, Teicher M, Abeles M (2006) Temporally precise cortical firing patterns are associated with distinct action segments. *J Neurophysiol* 96:2645–2652

67. Silver R, Boahen K, Grillner S, Kopell N, Olsen K (2007) Neurotech for neuroscience: unifying concepts, organizing principles, and emerging tools. *J Neurosci* 27:11,807–11,819
68. Smith LB, Yu C, Pereira AF (2011) Not your mothers view: the dynamics of toddler visual experience. *Dev Sci* 14(1):9–17
69. Song S, Miller KD, Abbott LF (2000) Competitive hebbian learning through spike-timing dependent synaptic plasticity. *Nat Neurosci* 3:919–926
70. Steels L (2003) Evolving grounded communication for robots. *Trends Cognit Sci* 7(7):308–312
71. Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language. A case study for colour. *Behav Brain Sci* 24(8):469–529
72. Steels L, Kaplan F, McIntyre A, Van Looveren J (2002) Crucial factors in the origins of word-meaning. In: Wray A (ed) *The transition to language*. Oxford University Press, Oxford, pp 252–271
73. Thelen E, Smith LB (1998) Dynamic systems theories. *Handbook of child psychology*
74. Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522
75. Tomasello M (2000) The item-based nature of childrens early syntactic development. *Trends Cognit Sci* 4(4):156–163
76. Tonkes B, Willes J (2002) Minimally biased learners and the emergence of language. In: Wray A (ed) *The transition to language*. Oxford University Press, Oxford
77. Twomey K, Morse A, Cangelosi A, Horst J (2014) Competition affects word learning in a developmental robotic system. In: 14th neural computation and psychology workshop
78. Vernon D (2014) *Artificial cognitive systems: a primer*. MIT Press, Cambridge
79. Vernon D, Metta G, Sandini G (2007) A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Trans Evolut Comput* 11(2):151–180
80. Warneken F, Chen F, Tomasello M (2006) Cooperative activities in young children and chimpanzees. *Child Dev* 77(3):640–663
81. Zhou X, Wang B (2004) Preschool childrens representation and understanding of written number symbols. *Early Child Dev Care* 174(3):253–266
82. Zukow-Goldring P, Arbib MA (2007) Affordances, effectivities, and assisted imitation: caregivers and the directing of attention. *Neurocomputing* 70(13):2181–2193