How Developmental Psychology and Robotics Complement Each Other

Brian Scassellati
MIT Artificial Intelligence Lab
545 Technology Square – Room 938
Cambridge, Massachusetts 02139
scaz@ai.mit.edu

Abstract

This paper presents two complementary ideas relating the study of human development and the construction of intelligent artifacts. First, the use of developmental models will be a critical requirement in the construction of robotic systems that can acquire a large repertoire of motor, perceptual, and cognitive capabilities. Second, robotic systems can be used as a test-bed for evaluating models of human development much in the same way that simulation studies are currently used to evaluate cognitive models. To further explore these ideas, two examples from the author's own work will be presented: the use of developmental models of hand-eye coordination to simplify the task of learning to reach for a visual target and the use of a humanoid robot to evaluate models of normal and abnormal social skill development.

Introduction

Research on human development and research on the construction of intelligent artifacts can and should be complementary. Studies of human development have produced a great variety of theories, models, and experimental constructs which have long been an inspiration for implementations of robotic systems. Research from human development has often served as the inspiration for both challenging research questions and useful task decompositions. However, computational studies of developmental processes have had little impact on the theoretical constructs present in developmental psychology today, and the influence of robotics on developmental studies has been almost completely absent. In this paper, I will argue that not only will robotics come to rely upon human development for inspiration and practical theories, but also will human development profit from the evaluation and experimentation opportunities that robotics offers. In next section, I will briefly describe the practical and theoretical ways in which developmental models aid in the construction of intelligent artifacts by focusing on the implementation of simple hand-eye coordination that our group has implemented on a humanoid robot. In the final section, I will discuss work in progress on using a robotic platform as a unique test-bed to evaluate models of social skill development for both normal and autistic individuals.

How Developmental Psychology Impacts Robotics

Developmental psychology is most typically employed in robotics research as a source of inspiration. Questions that have been addressed in the developmental psychology literature (such as the how infants learn to orient to salient stimuli and how children learn to navigate unfamiliar locations) have focused on issues that have also been of interest to the robotics community. Models from developmental psychology often offer behavioral decomposition and observations about task performance which may provide an outline for a software architecture. Techniques for studying skill progressions have also been adapted as evaluation techniques for robotics systems.

However, a developmental approach to robot construction also provides practical benefits. Human development exploits a gradual increase in both internal complexity (perceptual and motor) and external complexity (task and environmental complexity regulated by the instructor) to optimize the acquisition of new skills. For example, infants are born with low acuity vision which simplifies the visual input they must process. The infant's visual performance develops in step with their ability to process the influx of stimulation (Johnson). The same is true for the motor system. Newborn infants do not have independent control over each degree of freedom of their limbs, but through a gradual increase in the granularity of their motor control they learn to coordinate the full complexity of their bodies. A process in which the acuity of both sensory and motor systems are gradually increased significantly reduces the difficulty of the learning problem (Thelen & Smith). The caregiver also acts to gradually increase the task complexity by structuring and controlling the complexity of the environment. Our group has previously argued that developmental approaches to robot construction produce systems that can scale naturally to more complex tasks and problem domains by optimizing learning in a similar way (Brooks, (Ferrell), Irie, Kemp, Marjanović, Scassellati & Williamson). By ex-

Figure 1: Cog, an upper-torso humanoid robot with twenty-one degrees of freedom and a variety of sensory systems including visual, auditory, tactile, kinesthetic, and vestibular systems.

ploiting a gradual increase in complexity both internal and external, while reusing structures and information gained from previously learned behaviors, increasingly more sophisticated behaviors can be acquired (Ferrell; Scassellati).

Example #1 : Hand-Eye Coordination

Diamond (1990) has shown that infants between five and twelve months of age progress through a number of distinct phases in the development of visually guided reaching. In this progression, infants in later phases consistently demonstrate more sophisticated reaching strategies to retrieve a toy in more challenging scenarios. As the infant's reaching competency develops, later stages incrementally improve upon the competency afforded by the previous stages. Within our group, Marjanović, Scassellati & Williamson (1996) applied a similar bootstrapping technique to enable a humanoid robot (shown in Figure 1) to learn to point to a visual target. This pointing behavior is learned over many repeated trials without human supervision, using gradient descent methods to train forward and inverse mappings between a visual parameter space and an arm position parameter space. Without a developmental perspective, the problem of pointing to a visual target is a degenerate $R^2 \rightarrow R^4$ sensory-motor mapping problem with no obvious training signal; the position of the target in the visual coordinates (a two-dimensional quantity) must be converted into an arm trajectory for the four degrees of freedom in the arm. Using the behavioral decomposition Diamond (1990) observed in infants, Marjanović et al. (1996) reduced this $R^2 \rightarrow R^6$ function into a pair of $R^2 \rightarrow R^2$ learned functions and a fixed $R^2 \rightarrow R^4$ non-degenerate function with obvious error signals.

From an external perspective, the robot's behavior is quite rudimentary. Given a visual stimulus, typically by a researcher waving an object in front of its cameras, the robot saccades to foveate on the target, and then reaches out its arm toward the target. Early reaches are inaccurate, and often in the wrong direction altogether, but after a few hours of practice the accuracy improves drastically. To reach to a visual target, the robot must learn the mapping from the target's image coordinates $\vec{x} = (x, y)$ to the coordinates of the arm motors $\vec{\alpha} = (\alpha_0 ... \alpha_5)$ (see Figure 2). To achieve this, the robot first learns to foveate the target using a saccade map $\vec{S} : \vec{x} \rightarrow \vec{e}$ which relates positions in the camera image with the motor commands necessary to foveate the eye at that location ($\vec{e} = (pan, tilt)$). Once the target is foveated, the robot must learn a ballistic movement mapping head-centered coordinates $\vec{e}$ to arm-centered coordinates $\vec{\alpha}$. To simplify the dimensionality problems involved in controlling a six degree-of-freedom arm, arm positions are specified as a linear combination of basis posture primitives.

Both the saccade map and the ballistic arm map are constructed by on-line learning algorithms. The saccade map is trained using a correlation-based tracker. The error signal is a vector in image coordinates, and can be used to directly train the mapping. Once the saccade map has been trained, the ballistic map is trained using by comparing arm motor command signals with visual motion feedback clues to localize the arm in visual space (see Figure 3). By visually tracking the moving arm, we can obtain its final position in image coordinates. The vector from the tip of the arm in the image to the center of the image is the visual error signal, which can be converted into an error in gaze coordinates using the saccade mapping. In this way, the knowledge gained from learning to foveate a target transforms the ballistic arm error into an error signal that can be used to train the arm directly. This re-use allows the learning algorithms to operate continually, in real time, and in an unstructured "real-world" environment without using explicit world coordinates or complex kinematics. This technique successfully trains a reaching behavior within approximately three hours of self-supervised training. Video clips of Cog reaching to a visual target are available from `http://www.ai.mit.edu/projects/cog/`, and additional details on this method can be found in Marjanović et al. (1996).

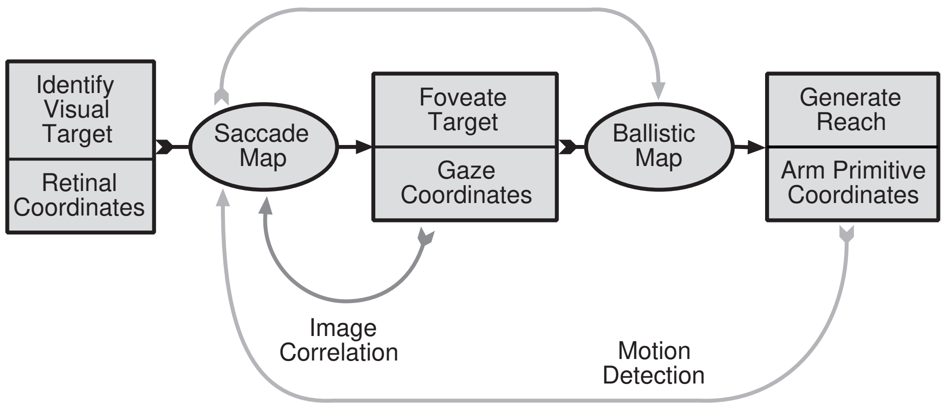

Figure 2: Reaching to a visual target is the product of two sub-skills: foveating a target and generating a ballistic reach from that eye position. Image correlation can be used to train a saccade map which transforms retinal coordinates into gaze coordinates (eye positions). This saccade map can then be used in conjunction with motion detection to train a ballistic map which transforms gaze coordinates into a ballistic reach.

How Robotics Can Impact Developmental Psychology

I have proposed that humanoid robotics research can also investigate scientific questions about the nature of human intelligence (Scassellati; Scassellati). Humanoid robots can serve as a unique tool to investigators in the cognitive sciences. Robotic implementations of cognitive, behavioral, and developmental models provide a test-bed for evaluating the predictive power and validity of those models. An implemented robotic model allows for more accurate testing and validation of these models through controlled, repeatable experiments. Slight experimental variations can be used to isolate and evaluate single factors (whether environmental or internal) independent of many of the confounds that affect normal behavioral observations. Experiments can also be repeated with nearly identical conditions to allow for easy validation. Further, internal model structures can be manipulated to observe the quantitative and qualitative effects on behavior. A robotic model can also be subjected to controversial testing that is potentially hazardous, costly, or unethical to conduct on humans; the "boundary conditions" of the models can be explored by testing alternative learning and environmental conditions. Finally, a robotic model can be used to suggest and evaluate potential intervention strategies before applying them to human subjects.

Example #2 : Development of Joint Reference

One of the critical precursors to social learning in human development is the ability to selectively attend to an object of mutual interest. Humans have a large repertoire of social cues, such as gaze direction, pointing gestures, and postural cues, that all indicate to an observer which object is currently under consideration. These abilities, collectively named mechanisms of joint (or shared) attention, are vital to the normal development of social skills in children. Joint attention to objects and events in the world serves as the initial mechanism for infants to share experiences with others and to negotiate shared meanings. Joint attention is also a mechanism for allowing infants to leverage the skills and knowledge of an adult caretaker in order to learn about their environment, in part by allowing the infant to manipulate the behavior of the caretaker and in part by providing a basis for more complex forms of social communication such as language and gestures.

Joint attention has been investigated by researchers in a variety of fields. Experts in child development are interested in these skills as part of the normal developmental course that infants acquire extremely rapidly, and in a stereotyped sequence (Scaife & Bruner; Moore & Dunham). Additional work on the etiology and behavioral manifestations of pervasive developmental disorders such as autism and Asperger's syndrome have focused on disruptions to joint attention mechanisms and demonstrated how vital these skills are in human social interactions (Cohen & Volkmar; Baron-Cohen). Philosophers have been interested in joint attention both as an explanation for issues of contextual grounding and as a precursor to a theory of other minds (Whiten; Dennett). Evolutionary psychologists and primatologists have focused on the evolution of these simple social skills throughout the animal kingdom as a means of evaluating both the presence of theory of mind and as a measure of social functioning (Povinelli & Preuss; Hauser; Premack).

The inspiration for an implementation of joint reference comes from Baron-Cohen (1995). Baron-Cohen's model gives a coherent account of the observed developmental stages of joint attention behaviors in both normal and blind children, the observed deficiencies in joint

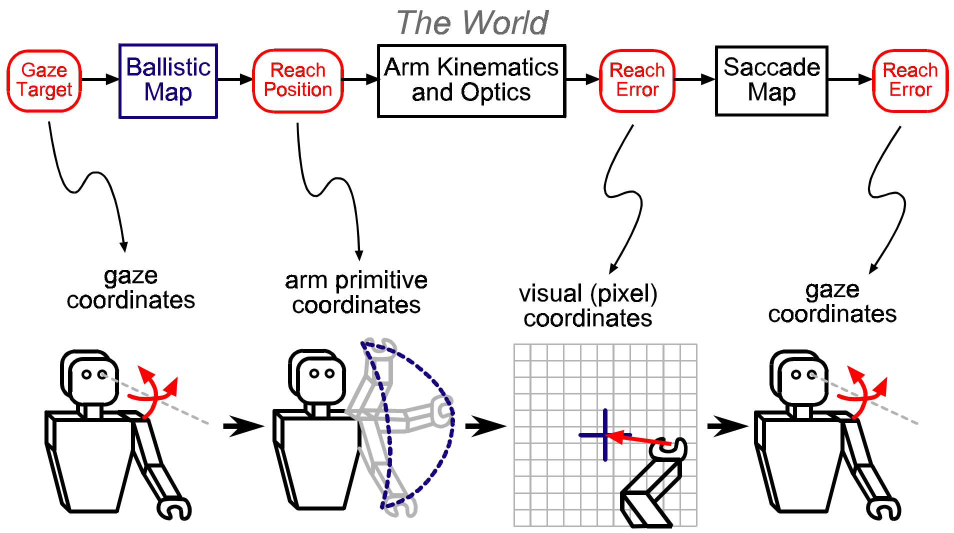

Figure 3: Generation of error signals from a single reaching trial. Once a visual target is foveated, the gaze coordinates are transformed into a ballistic reach by the ballistic map. By observing the position of the moving hand, we can obtain a reaching error signal in image coordinates, which can be converted back into gaze coordinates using the saccade map.

attention of children with autism, and a partial explanation of the observed abilities of primates on joint attention tasks. The model provides both a skill decomposition and a potential system architecture for constructing a system that can recognize and respond to eye contact, gaze direction, and imperative and declarative pointing gestures.

A robotic implementation of Baron-Cohen's model is currently under construction (Scassellati). We have already implemented a perceptual system capable of finding faces and eyes (Scassellati). The system first locates potential face locations in the peripheral image using a template-based matching algorithm. Once a potential face location has been identified, the robot saccades to that target using the saccade mapping S described earlier. The location of the face in peripheral image coordinates ($p_{(x,y)}$) is then mapped into foveal image coordinates ($f_{(x,y)}$) using a second learned mapping, the foveal map $F: p_{(x,y)} \mapsto f_{(x,y)}$. The location of the face within the peripheral image can then be used to extract the sub-image containing the eye for further processing. This technique has been successful at locating and extracting sub-images that contain eyes under a variety of conditions and from many different individuals. Additional modules including a context-sensitive attention system (Breazeal & Scassellati), a system of human-like eye and neck movement (Brooks, Breazeal, Marjanovic, Scassellati & Williamson), and a system for regulating interaction intensities (Breazeal & Scassellati) have also been implemented.

Advantages of a Robotic Implementation A robotic approach to studies of joint attention and social skill development has three main advantages. First, human observers readily anthropomorphize their social interactions with a human-like robot. Second, the construction of a physically embodied system may be computationally simpler than the construction of a simulation of sufficient detail. Third, the skills that must be implemented to test these models are useful for a variety of other practical robotics tasks.

Interactions with a robotic agent are easily anthropomorphized by children and adults. An embodied system with human form allows for natural social interactions to occur without any additional training or prompting. Observers need not be trained in special procedures necessary to interact with the robot; the same behaviors that they use for interacting with other people allow them to interact naturally with the robot. In our experience, and in the empirical studies by Reeves & Nass (1996), people readily treat a robot as if it were another person. Human form also provides important task constraints on the behavior of the robot. For example, to observe an object carefully, our robot must orient its head and eyes toward a target. These task constraints allow observers to easily interpret the behavior of the robot.

A second reason for choosing a robotic implementation is that physical embodiment may actually simplify the computation necessary for this task. The direct physical coupling between action and perception reduces the need for an intermediary representation. For an embodied system, internal representations can be ultimately grounded in sensory-motor interactions with the world (Lakoff); there is no need to model aspects of the environment that can simply be experi-

Figure 4: Examples of successful face and eye detections. The system locates faces in the peripheral camera, saccades to that position, and then extracts the eye image from the foveal camera. The position of the eye is inexact, in part because the human subjects are not motionless.

enced (Brooks; Brooks). The effects of gravity, friction, and natural human interaction are obtained for free, without any computation. Embodied systems can also perform some complex tasks in relatively simple ways by exploiting the properties of the complete system. For example, when putting a jug of milk in the refrigerator, you can exploit the pendulum action of your arm to move the milk (Greene). The swing of the jug does not need to be explicitly planned or controlled, since it is the natural behavior of the system. Instead of having to plan the whole motion, the system only has to modulate, guide and correct the natural dynamics.

Third, the social skills that we must implement to test these models are important from an engineering perspective. A robotic system that can recognize and engage in joint attention behaviors will allow for human-machine interactions that have previously not been possible. The robot would be capable of learning from an observer using normal social signals in the same way that human infants learn; no specialized training of the observer would be necessary. The robot would also be capable of expressing its internal state (emotions, desires, goals, etc.) through social interactions without relying upon an artificial vocabulary. Further, a robot that can recognize the goals and desires of others will allow for systems that can more accurately react to the emotional, attentional, and cognitive states of the observer, can learn to anticipate the reactions of the observer, and can modify its own behavior accordingly.

Implementing this progression for a robotic system provides a simple means of bootstrapping behaviors. The capabilities used in detecting and maintaining eye contact can be extended to provide a rough angle of gaze. By tracking along this angle of gaze, and watching for objects that have salient color, intensity, or motion, our robot can mimic the ecological strategy. From an ecological mechanism, we can refine the algorithms for determining gaze and add mechanisms for determining vergence. A rough geometric strategy can then be implemented, and later refined through feedback from the caretaker. A representational strategy requires the ability to maintain information on salient objects that are outside of the field of view including information on their appearance, location, size, and salient properties. The implementation of this strategy requires us to make assumptions about the important properties of objects that must be included in a representational structure, a topic beyond the scope of this paper.

Evaluating the Robotic Implementation A robotic implementation of a behavioral model provides a standardized evaluation mechanism. Behavioral observation and classification techniques that are used on children and adults can be applied to the behavior of our robot with only minimal modifications. Because of their use in the diagnosis and assessment of autism and related disorders, evaluation tools for joint attention mechanisms, such as the Vineland Adaptive Behavior Scales, the Autism Diagnostic Interview, and the Autism Diagnostic Observation Schedule, have been extensively studied (Sparrow, Marans, Klin, Carter, Volkmar & Cohen; Powers). With the evaluations obtained from these tools, the success of our implementation efforts can be tested using the same criteria that are applied to human behaviors. The behavior of the complete robotic implementation can be compared with developmental data from normal children. Furthermore, by inhibiting specific modules within the model, the robot should produce behavior that can be compared with developmental data from autistic children. With these evaluation techniques, we can determine the extent to which our model matches the observed biological data. However, what conclusions can we draw from the outcomes of these studies?

One possible outcome is that our robotic implementation will match the expected behavior evaluations, that is, the complete system will demonstrate normal uses of joint attention. In this case, our efforts have provided evidence that the model is internally consistent in producing the desired behaviors, but says nothing about the underlying biological processes. We can verify that the model provides a possible explanation for the normal (and abnormal) development of joint attention, but we cannot verify that this model accurately reflects what happens in biology.

If the robotic implementation does not meet the same behavioral criteria, the reasons for the failure are significant. The implementation may be unsuccessful because of an internal logical flaw in the model. In this case, we can identify shortcomings of the proposed model and potentially suggest alternate solutions. A more difficult failure may result if our environmental conditions differ too significantly from normal human social interactions. While the work of Reeves & Nass (1996) leads us to believe that this result will not occur, this possibility allows us to draw conclusions only about our implementation and not the model or the underlying biological factors.

Future Work The implementation of Baron-Cohen's model is still work in progress. All of the basic sensory-motor skills have been demonstrated. The robot can move its eyes in many human-like ways, including saccades, vergence, tracking, and maintaining fixation through vestibulo-ocular and opto-kinetic reflexes. Orientation with the neck to maximize eye range has been implemented, as well as coordinated arm pointing. Perceptual components of eye detection have also been constructed; the robot can detect and foveate faces to obtain high-resolution images of eyes.

These initial results are incomplete, but have provided encouraging evidence that the technical problems faced by an implementation of this nature are within our grasp. Cog's perceptual systems have been successful at finding faces and eyes in real-time, and in real-world environments. Simple social behaviors, such as eye-neck orientation and head-nod imitation, have been easy to interpret by human observers who have found their interactions with the robot to be both believable and entertaining.

Our future work will focus on the construction and implementation of the remainder of the modules from Baron-Cohen's model. From an engineering perspective, this approach has already succeeded in providing adaptive solutions to classical problems in behavior integration, space-variant perception, and the integration of multiple sensory and motor modalities. From a scientific perspective, we are optimistic that when completed, this implementation will provide new insights and evaluation methods for models of social development.

Acknowledgements

Parts of this research are funded by DARPA/ITO under contract number DABT 63-99-1-0012 and parts have been funded by ONR under contract number N00014-95-1-0600, "A Trainable Modular Vision System." Many of the results included in this paper have been previously reported (Scassellati; Scassellati; Brooks, Breazeal, Marjanovic, Scassellati & Williamson; Marjanović et al.; Brooks, (Ferrell), Irie, Kemp, Marjanović, Scassellati & Williamson).

References

Baron-Cohen, S. (1995), *Mindblindness*, MIT Press.

Breazeal, C. & Scassellati, B. (1999), A context-dependent attention system for a social robot, *in* '1999 International Joint Conference on Artificial Intelligence'.

Breazeal, C. & Scassellati, B. (2000), 'Infant-like Social Interactions between a Robot and a Human Caretaker', *Adaptive Behavior*. To appear.

Brooks, R. A. (1986), 'A Robust Layered Control System for a Mobile Robot', *IEEE Journal of Robotics and Automation* **RA-2**, 14–23.

Brooks, R. A. (1991), 'Intelligence Without Representation', *Artificial Intelligence Journal* **47**, 139–160. originally appeared as MIT AI Memo 899 in May 1986.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B. & Williamson, M. M. (1999), The Cog Project: Building a Humanoid Robot, *in* C. L. Nehaniv, ed., 'Computation for Metaphors, Analogy and Agents', Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*, Springer-Verlag.

Brooks, R. A., (Ferrell), C. B., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B. & Williamson, M. M. (1998), Alternative Essences of Intelligence, *in* 'Proceedings of the American Association of Artificial Intelligence (AAAI-98)'.

Cohen, D. J. & Volkmar, F. R., eds (1997), *Handbook of Autism and Pervasive Developmental Disorders*, second edn, John Wiley & Sons, Inc.

Dennett, D. C. (1991), *Consciousness Explained*, Little, Brown, & Company.

Diamond, A. (1990), Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of Inhibitory Control in Reaching, *in* 'The Development and Neural Bases of Higher Cognitive Functions', Vol. 608, New York Academy of Sciences, pp. 637–676.

Ferrell, C. (1996), Orientation Behavior using Registered Topographic Maps, *in* 'From Animals to Animats: Proceedings of 1996 Society of Adaptive Behavior', Cape Cod, Massachusetts, pp. 94–103.

Greene, P. H. (1982), 'Why is it easy to control your arms?', *Journal of Motor Behavior* **14**(4), 260–286.

Hauser, M. D. (1996), *Evolution of Communication*, MIT Press.

Johnson, M. H. (1993), Constraints on Cortical Plasticity, *in* M. H. Johnson, ed., 'Brain Development and Cognition: A Reader', Blackwell, Oxford, pp. 703–721.

Lakoff, G. (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago, Illinois.

Marjanović, M. J., Scassellati, B. & Williamson, M. M. (1996), Self-Taught Visually-Guided Pointing for a Humanoid Robot, *in* 'From Animals to Animats: Proceedings of 1996 Society of Adaptive Behavior', Cape Cod, Massachusetts, pp. 35–44.

Moore, C. & Dunham, P. J., eds (1995), *Joint Attention: Its Origins and Role in Development*, Erlbaum.

Povinelli, D. J. & Preuss, T. M. (1995), 'Theory of Mind: evolutionary history of a cognitive specialization', *Trends in Neuroscience*.

Powers, M. D. (1997), Behavioral Assessment of Individuals with Autism, *in* Cohen & Volkmar (1997).

Premack, D. (1988), "Does the chimpanzee have a theory of mind?" revisited, *in* R. Byrne & A. Whiten, eds, 'Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans.', Oxford University Press.

Reeves, B. & Nass, C. (1996), *The media equation : how people treat computers, televisions, and new media like real people and places*, Cambridge University Press.

Scaife, M. & Bruner, J. (1975), 'The capacity for joint visual attention in the infant.', *Nature* **253**, 265–266.

Scassellati, B. (1996), Mechanisms of Shared Attention for a Humanoid Robot, *in* 'Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium', AAAI Press.

Scassellati, B. (1998*a*), Building Behaviors Developmentally: A New Formalism, *in* 'Integrating Robotics Research: Papers from the 1998 AAAI Spring Symposium', AAAI Press.

Scassellati, B. (1998*b*), Finding Eyes and Faces with a Foveated Vision System, *in* 'Proceedings of the American Association of Artificial Intelligence (AAAI-98)'.

Scassellati, B. (1999), Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot, *in* C. L. Nehaniv, ed., 'Computation for Metaphors, Analogy and Agents', Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*, Springer-Verlag.

Sparrow, S., Marans, W., Klin, A., Carter, A., Volkmar, F. R. & Cohen, D. J. (1997), Developmentally Based Assessments, *in* Cohen & Volkmar (1997).

Thelen, E. & Smith, L. (1994), *A Dynamic Systems Approach to the Development of Cognition and Action*, MIT Press, Cambridge, MA.

Whiten, A., ed. (1991), *Natural Theories of Mind*, Blackwell.