Investigating Models of Social Development Using a Humanoid Robot

(Invited Paper)

Brian Scassellati
Yale University
Department of Computer Science
51 Prospect Street
New Haven, CT 06520
Email: scaz@cs.yale.edu

***Abstract*— Human social dynamics rely upon the ability to correctly attribute beliefs, goals, and percepts to other people. The set of abilities that allow an individual to infer these hidden mental states based on observed actions and behavior has been called a "theory of mind" [1]. Drawing from the models of Baron-Cohen [2] and Leslie [3], a novel architecture called *embodied theory of mind* was developed to link high-level cognitive skills to the low-level perceptual abilities of a humanoid robot. The implemented system determines visual saliency based on inherent object attributes, high-level task constraints, and the attentional states of others. Objects of interest are tracked in real-time to produce motion trajectories which are analyzed by a set of naive physical laws designed to discriminate animate from inanimate movement. Animate objects can be the source of attentional states (detected by finding faces and head orientation) as well as intentional states (determined by motion trajectories between objects). Individual components are evaluated by comparisons to human performance on similar tasks, and the complete system is evaluated in the context of a basic social learning mechanism that allows the robot to mimic observed movements.**

I. Introduction

The term "theory of mind" has been used to identify a collection of socially-mediated skills which are useful in relating the individual's behavior within a social context. Examples of these skills include detecting eye contact, recognizing what someone else is looking at, pointing to direct attention to interesting objects, and understanding that other people have ideas that differ from one's own. Research from many different disciplines has focused on theory of mind. Students of philosophy have been interested in the understanding of other minds and the representation of knowledge in others [4]. Ethologists have also focused on the presence (and absence) of these social skills in primates and other animals [5]–[7]. Research on the development of social skills in children has focused on characterizing the developmental progression of social abilities [8]–[10] and on how these skills result in conceptual changes and the representational capacities of infants [11], [12]. Furthermore, research on pervasive developmental disorders such as autism has focused on the selective impairment of these social skills [13]–[15].

This paper presents two popular and influential models [2], [3], which attempt to link together multi-disciplinary research into a coherent developmental explanation. We then discuss the implications of these models for the construction of humanoid robots that engage in natural human social dynamics and will also highlight some of the issues involved in implementing the structures that these models propose. Finally, we will describe a hybrid model called *embodied theory of mind* that links together ideas from both Baron-Cohen and Leslie with a grounded perceptual system. The hybrid model was implemented on a humanoid robot and evaluated using a simple social learning scenario.

II. Leslie's Model

Leslie's [3] theory treats the representation of causal events as a central organizing principle to theories of object mechanics and theories of other minds much in the same way that the notion of number may be central to object representation. According to Leslie, the world is naturally decomposed into three classes of events based upon their causal structure; one class for *mechanical agency*, one for *actional agency*, and one for *attitudinal agency*. Leslie argues that evolution has produced independent domain-specific modules to deal with each of these classes of event.

The Theory of Body module (ToBY) deals with events that are best described by mechanical agency, that is, they can be explained by the rules of *mechanics*. ToBY's goal is to describe the world in terms of the mechanics of physical objects and the events they enter into. ToBY in humans is believed to operate on two types of visual input: a three-dimensional object-centered representation from high level cognitive and visual systems and a simpler motion-based system. This motion-based system accounts for the causal explanations that adults give (and the causal expectations of children) to the "billiard ball" type launching displays pioneered by Michotte [16].

ToBY is followed developmentally by the emergence of a Theory of Mind Mechanism (ToMM) which develops in two phases, which I will denote ToMM-1 and ToMM-2 after [2]. Just as ToBY deals with the physical laws that govern objects, ToMM deals with the psychological laws that govern agents. ToMM-1 explains events in terms of the intent and goals of agents, that is, their *actions*. The primitive representations of actions such as approach, avoidance, and escape are constructed by ToMM-1. This system of detecting goals and actions begins to emerge at around 6 months of age, and is most often characterized by attention to eye gaze. ToMM-2 explains events in terms of the *attitudes* and beliefs of agents; it deals with the representations of beliefs and how mental states can drive behavior relative to a goal. This system develops

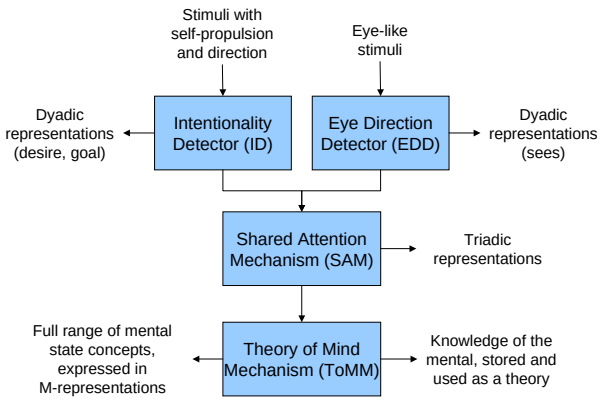

Fig. 1. Block diagram of Baron-Cohen's model of the development of theory of mind. See text for description. Adapted from Baron-Cohen [2].

gradually, with the first signs of development beginning between 18 and 24 months of age and completing sometime near 48 months. ToMM-2 employs the M-representation, a meta-representation which allows truth properties of a statement to be based on mental states rather than observable stimuli. ToMM-2 is a required system for understanding that others hold beliefs that differ from our own knowledge or from the observable world, for understanding different perceptual perspectives, and for understanding pretense and pretending.

III. Baron-Cohen's Model

While Leslie's model has a clean conceptual division of the external world into three spheres of causality, Baron-Cohen's model is more easily grounded in perceptual processes. Baron-Cohen's model assumes two forms of perceptual information are available as input. The first percept describes all stimuli in the visual, auditory, and tactile perceptual spheres that have self-propelled motion. The second percept describes all visual stimuli that have eye-like shapes. Baron-Cohen proposes that the set of precursors to a theory of mind, which he calls the "mindreading system," can be decomposed into four distinct modules (see figure 1).

The first module interprets self-propelled motion of stimuli in terms of the primitive volitional mental states of goal and desire. This module, called the intentionality detector (ID) produces dyadic representations that describe the basic movements of approach and avoidance. For example, ID can produce representations such as "he wants the food" or "she wants to go over there". This module only operates on stimuli that have self-propelled motion, and thus pass a criteria for distinguishing stimuli that are potentially animate (agents) from those that are not (objects). Baron-Cohen speculates that ID is a part of the innate endowment that infants are born with.

The second module processes visual stimuli that are eye-like to determine the direction of gaze. This module, called the eye direction detector (EDD), has three basic functions. First, it detects the presence of eye-like stimuli in the visual field. Human infants have a preference to look at human faces, and spend more time gazing at the eyes than at other parts of the face. Second, EDD computes whether the eyes are looking at it or at something else. Baron-Cohen proposes that having someone else make eye contact is a natural psychological releaser that produces pleasure in human infants (but may produce more negative arousal in other animals). Third, EDD interprets gaze direction as a perceptual state, that is, EDD codes dyadic representational states of the form "agent sees me" and "agent looking-at not-me".

The third module, the shared attention mechanism (SAM), takes the dyadic representations from ID and EDD and produces triadic representations of the form "John sees (I see the girl)". Embedded within this representation is a specification that the external agent and the self are both attending to the same perceptual object or event. This shared attentional state results from an embedding of one dyadic representation within another. SAM additionally can make the output of ID available to EDD, allowing the interpretation of eye direction as a goal state. By allowing the agent to interpret the gaze of others as intentions, SAM provides a mechanism for creating nested representations of the form "John sees (I want the toy)".

The last module, the theory of mind mechanism (ToMM), provides a way of representing epistemic mental states in other agents and a mechanism for tying together our knowledge of mental states into a coherent whole as a usable theory. ToMM first allows the construction of representations of the form "John believes (it is raining)". ToMM allows the suspension of the normal truth relations of propositions (referential opacity), which provides a means for representing knowledge states that are neither necessarily true nor match the knowledge of the organism, such as "John thinks (Elvis is alive)". Baron-Cohen proposes that the triadic representations of SAM are converted through experience into the M-representations of ToMM.

IV. Implications for Humanoid Robots

The most exciting aspect of these models from an engineering perspective is that they attempt to describe the perceptual and motor skills that serve as precursors to the more complex theory of mind capabilities. These decompositions serve as an inspiration and a guideline for building robotic systems that can engage in complex social interactions; they provide a much-needed division of a rather ambiguous ability into a set of observable, testable predictions about behavior. While it cannot be claimed with certainty that following the outlines that these models provide will produce a robot that has the same abilities, the evolutionary and developmental evidence for this skill decomposition does give us hope that these abilities are critical elements of the larger goal. Additionally, the grounding of high-level perceptual abilities to observable sensory and motor capabilities provides an evaluation mechanism for measuring the amount of progress that is being made. Robotic implementations of these systems can be evaluated using the same behavioral and observational metrics that are

used to assess the presence or absence of that same skill in children.

Perhaps more importantly, the theory of mind models are interesting from a theoretical standpoint in that they serve as a bridge between skills that are often thought to be high-level cognitive phenomena and low-level skills that are strongly perceptual processes. This link allows for a bottom-up engineering approach to begin to address questions about high-level cognitive tasks by showing how these tasks can be grounded into perceptual and motor capabilities. While this connection may seem obvious given the psychological data, it is often difficult in fields (including robotics) that are driven primarily by bottom-up design to see how these low-level abilities might someday scale to more complex questions. Similarly, in fields (including much of classical artificial intelligence) where top-down design is the status quo, it is difficult to bind abstract reasoning to realistic sensory data. Bottom-up design tends to result in systems that are robust and practical, but that in many ways fail to construct interesting and complex behavior. Top-down design will often result in systems that are elegant abstractions, but that have little hope of being usable in a real system. These models of theory of mind provide the insight to construct a system that is truly grounded in the real-world sensory and motor behaviors but that also can begin to engage some interesting high-level cognitive questions.

From a robotics standpoint, the most salient differences between the two models are the ways in which they divide perceptual tasks. Leslie cleanly divides the perceptual world into animate and inanimate spheres and allows for further processing to occur specifically to each type of stimulus. Baron-Cohen does not divide the perceptual world quite so cleanly but does provide more detail on limiting the specific perceptual inputs that each module requires. In practice, both models require remarkably similar perceptual systems (which is not surprising, since the behavioral data is not under debate). However, each perspective is useful in its own way in building a robotic implementation. At one level, the robot must distinguish between object stimuli that are to be interpreted according to physical laws and agent stimuli that are to be interpreted according to psychological laws. However, the specifications that Baron-Cohen provides will be necessary for building visual routines that have limited scope.

The high-level abstract representations postulated by each model also have implications for robotics. Leslie's model has a very elegant decomposition into three distinct areas of influence, but the interactions between these levels are not well specified. Connections between modules in Baron-Cohen's model are better specified, but they are still less than ideal for a robotics implementation. Additionally, issues on how stimuli are to be divided between the competencies of different modules must be resolved for both models.

V. An Embodied Theory of Mind

Drawing from both Baron-Cohen's model and Leslie's model, we propose a hybrid architecture called the *embodied theory of mind*. This model connects modules similar to Leslie's ToBY and Baron-Cohen's EDD, ID, and SAM together with real perceptual processes and with links to physical behaviors. Because both Baron-Cohen and Leslie seek to explain the same underlying data, there is a great deal of overlap in the two representational systems. Leslie's ToMM-1 and ToMM-2 system overlap with the abilities of Baron-Cohen's EDD, ID, SAM, and ToMM modules. However, the emphasis that Leslie places on the theory of body module (ToBY) appears only as an input assumption to Baron-Cohen's model. The embodied theory of mind exploits these overlaps and extends the current models to behavior selection, attention, and more complex behavioral forms.

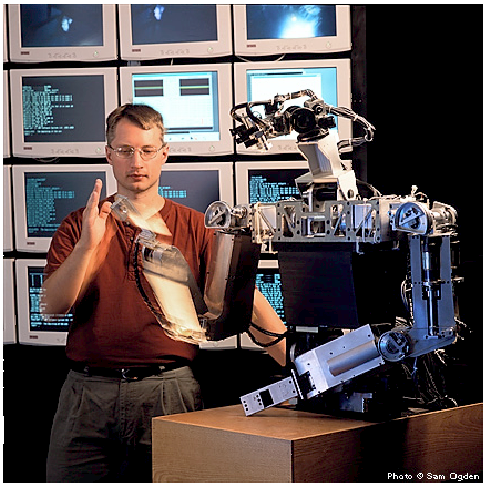

Fig. 2. Cog is an upper-torso humanoid robot with 22 degrees of freedom and a variety of sensory systems including visual, auditory, tactile, and kinesthetic sensing. Cog was designed to respond to natural social cues from a human instructor.

The humanoid robot called Cog served as the primary testbed for this research (see [17] for information on this platform). Cog is an upper-torso robot with 22 degrees of freedom and a variety of sensory systems including a binocular, foveated visual system, as well as auditory, tactile, and kinesthetic sensing (see Figure 2).

The primary insight in linking the two existing models together is that the theory of body module can act as a classifier for distinguishing self-propelled stimuli. The physical causal laws that ToBY encapsulates are really descriptions of how inanimate objects move through the world. ToBY can be transformed into a classifier by making the assumption that objects that are inanimate must obey these physical laws while objects that are animate will often break them. With this insight, we can begin to sketch out the connections between these modules (see figure 3). Visual input will be processed to form motion trajectories, similar to the trajectories observed in Michotte's experiments. These visual trajectories will then be analyzed by a set of naive physical laws in the theory of body module (ToBY). Objects that obey the laws of mechanical causality will be considered to be inanimate, while those that break mechanical causality laws will be classified as animate.

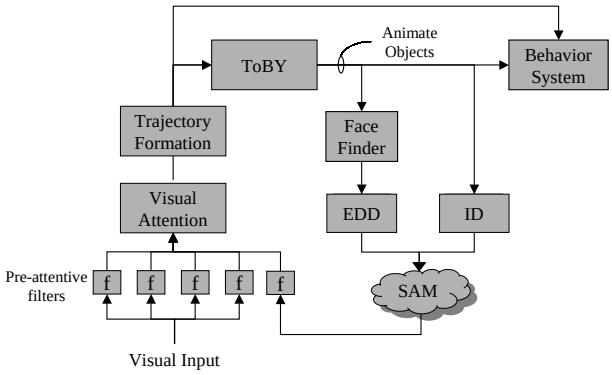

Fig. 3. Overview of the hybrid theory of mind model.

Baron-Cohen's model requires two types of input stimuli: objects with self-propelled motion and face-like objects. Animate stimuli trajectories serve directly as the input to Baron-Cohen's intentionality detector (ID). These animate trajectories will also then be processed by additional levels of image processing to find locations that contain faces. These face locations will then be the input to the eye direction detector module (EDD), which then feeds directly to the shared attention mechanism (SAM).

Connecting this rough outline to real perceptual systems and real motor response systems involves slightly more detail but still follows the same general principles. Raw visual input is processed by a number of low-level feature detectors (such as color, motion, and skin tone) which pre-attentively pick out areas of interest. These low-level filters will be combined with high-level task constraints and a habituation mechanism to select the most salient object in the scene. The attention system performs this selection and then directs limited computational and motor resources to the object of interest. The trajectories of interesting objects are tracked through time. These trajectories serve as the input to the theory of body mechanism, which employs an agent-based architecture to model the collective knowledge of many simple rules of naive physics. Any objects that violate the naive physical laws are declared animate and are subject to further processing by the initial modules of Baron-Cohen's model. Animate stimuli are processed by a multi-stage face detection system. Any faces in the scene attract the attention of the robot, which then uses a sequence of post-attentive processing steps to determine the orientation of the individual. These perceptual systems directly drive behaviors including head orientation, gaze direction, and pointing gestures. In addition, a simple social learning system has been implemented to demonstrate the effects of these social cues on imitative learning. Animate trajectories are processed by a simple intentionality detector that picks out relationships between animate objects and other objects based on a simple representation of approach and avoidance. These two representations trigger shared attention behaviors by applying an additional measurement of object saliency based on the attentional and intentional state of the observed individual.

Complete details on the implementation can be obtained from [18], [19]. In this paper, we focus on the implementation of joint reference as an example of the usefulness of this type of modeling.

A. Implementing Joint Reference

In the model of Baron-Cohen [2], the shared attention mechanism (SAM) links an attentional state to behavior that directs the robot's attention. In Baron-Cohen's terms, SAM is a "neurocognitive mechanism" rather than a module in sense of Fodor [8]. However, the treatment of SAM has always been as a distinct modular component – encapsulated knowledge that can be selectively present or absent. In the implementation discussed here, joint reference is not explicitly represented as a modular component. Rather, it is a property of a feedback mechanism between the head pose detection system and the attention system. This feedback loop, combined with the existing behavioral systems, produces the same joint reference behaviors as would be generated by SAM.

To complete the feedback between the perceptual processes that detect salient social cues and the behavioral systems that produce attentive behavior, a simple transformation must be employed.[1] The output of the head pose detection system is a data structure that includes the location of the face, the scale of the face, and the orientation of the head in terms of yaw, pitch, and roll. The inputs to the attention system are all structured in terms of retinotopic maps. The area of attention is mapped to the retinotopic input maps using a cone that originates at the center of the face location and extends along an angle that matches the projection of the head orientation. The intensity of the cone is at a maximum at its origin and degrades by 10% every fifteen pixels of distance from the origin. This gives both a directional differential and a distance differential which biases the robot to attend to the first salient object along that scan path. In practice, a cone with an extent of 15 degrees to either side of the orientation angle was found to be effective.

The addition of a joint reference input to the attention system is not a capability originally envisioned by Wolfe [21]. While there is little evidence that these joint reference behaviors are at the same perceptual level as the other pre-attentive filters in human visual behavior, this implementation choice is a simple method to allow all of the robust behaviors that had previously been designed to act on the output of attentional processes to be driven by joint reference without the introduction of any additional mechanisms. The relative influence of joint reference can easily be modified simply by changing the weighting that is applied to that input channel in the attentional process.

In addition to driving attentional responses such as orientation and pointing behaviors, the effect of joint reference can also be applied to select appropriate trajectories to mimic.

[1]By modifying the fidelity of this transformation, the first three of Butterworth's [20] stages of joint reference development can be achieved, although due to perceptual limitations only the first two were successfully demonstrated on the robot. For a full discussion, see [19].

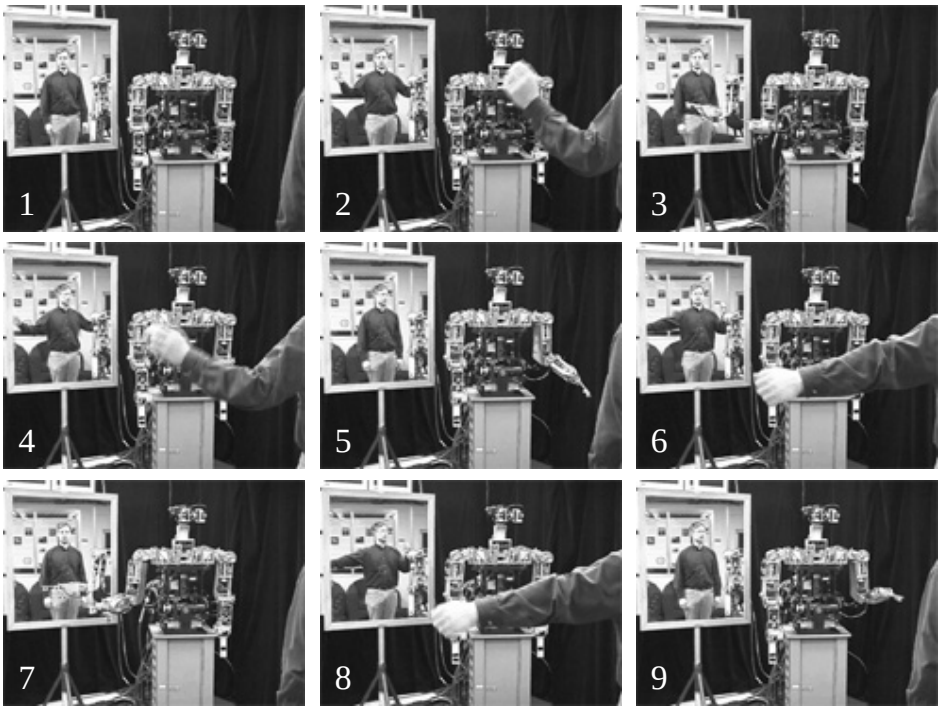

Fig. 4. Nine frames from a video sequence showing the application of joint reference for selection of trajectories for mimicry. In this video, a large mirror was positioned behind the robot, outside its field of view, to permit the video camera to record both the actions of the human and the robot. When the human looks to the left and makes two arm movements (images 1-2), the robot responds by selecting an arm movement that matches the head orientation (image 3). Similarly, when the human looks to the right (image 4), the trajectory to the right becomes more salient, and the robot acts upon it by moving its left arm (image 5). Images 6-9 show the same effect for two arm movements that differ from each other. Approximately two seconds of video separated each of these images.

People tend to pay close attention to their movements and manipulations of objects. When attempting to instruct another individual, this tendency is even more pronounced. In this way, attention acts as a natural saliency cue by pointing out the important aspects of the social scene. On Cog, the integration of the joint reference cues into the attention system allows for the selection of salient trajectories based on joint reference to be implemented without any further software. Figure 4 shows an example of the influence of head orientation on mimicry. To allow both the robot's behavior and the human's behavior to be captured using only a single video camera, a large mirror was placed behind the robot. The robot could neither see nor reach the mirror. The human instructor then made either identical movements with both arms (images 1-5) or different movements with both arms (images 6-9) while looking and orienting either toward his own left (images 1-3 and 6-7) or right (images 4-5 and 8-9). To allow an easily observable behavioral difference, the robot was programmed to respond either with its left or right arm, depending on whether the robot selected a trajectory that was to the right or the left of a detected face. (Note that to act like a mirror image reflection, when the human acts with his left hand, the robot must respond with its right hand.) As figure 4 demonstrates, the response of the robot to joint reference cues can easily be reflected in the mimicry behavior.

One of the primary differences between the embodied theory of mind presented here and the original work of Baron-Cohen [2] is that the role of joint reference is not encapsulated within a single modular structure. The model presented here should not be taken as any sort of proof that the human system operates in the same way. It does however provide an existence proof that joint reference behavior can be produced without the need for a complex, encapsulated module. The embodied model provides a useful interface to behavior selection and can account for many of the basic properties observed in the development of joint reference skills in infants. This perspective is not unheard of within the developmental science community. In fact, shortly before his death, Butterworth [22] had begun to articulate a position that joint attention is based on the properties of system embodiment. Butterworth noted that aspects of the design of the human body allowed the social cues that indicate attentional states to be more easily perceived. We agree with Butterworth that joint reference is supported by the basic facts of embodiment and that it can be grounded in perceptual states without resorting to wholly cognitive explanations of behaviors.

VI. Conclusion

Based on the models of Baron-Cohen [2] and Leslie [3], we have proposed a hybrid model of the foundational skills for a theory of mind. This model, which we have called

the *embodied theory of mind*, grounds concepts that have traditionally been thought to be high-level cognitive properties (such as animacy and intent) to low-level perceptual properties. All aspects of the model were implemented on a complex humanoid robot to operate in natural environments and at interactive rates. The implemented model featured the following components:

- An attentional mechanism which combined low-level feature detectors (such as color saturation, motion, and skin color filters) with high-level motivational influences to select regions of interest.
- A "theory of body" module which determined whether an object was animate or inanimate based on a set of naive physical laws that operated solely on the spatial and temporal properties of the object's movement.
- An active sensorimotor system that detected faces at a large variety of scales using a color pre-filter and two shape-based metrics. This system also identified three features (the two eyes and the mouth) and used those features to determine the orientation of the person's head. This information on the attentional state of the observed person was then used to engage in joint reference behaviors, directing the robot's attention to the same object that the person was considering.
- A simple mechanism for detecting the basic intentional states of approach/desire and avoidance/fear. These classifications were determined by considering pairs of trajectories and allowing attributions of intent to only be applied to animate agents.

Individual components were evaluated by comparison with human judgments on similar problems and the complete system was evaluated in the context of social learning. A basic mimicry behavior was implemented by mapping a visual trajectory to a movement trajectory for one of Cog's arms. Both the mimicry behavior and behaviors that generated an attentional reference (pointing and head orientation) were made socially relevant by limiting responses to animate trajectories, by acting on objects that became salient through joint reference, and by acting on objects that were involved in an intentional relationship. This set of simple behaviors made a first step toward constructing a system that can use natural human social cues to learn from a naive instructor.

Although no claims have been made that this implementation reflects the kinds of processing that occurs in either humans or other animals, systems like this one represent a new kind of tool in the evaluation and testing of human cognitive models [17], [23]. In particular, this implementation is an existence proof for building joint reference behaviors without an explicit, encapsulated module. The implementation has also demonstrated a useful addition to Wolfe's Guided Search model by incorporating both habituation effects and the effects of joint reference. Furthermore, the implemented system gives an example of how to perceptually ground animacy and intentionality judgments in real perceptual streams.

ACKNOWLEDGMENT

This research was conducted at the MIT Artificial Intelligence Laboratory. Portions of this research were funded by DARPA/ITO under contract number DABT 63-99-1-0012, "Natural tasking of robots based on human interaction cues," and in part by an ONR/ARPA Vision MURI Grant (No. N00014-95-1-0600). The author is indebted to the members of the humanoid robotics group at MIT, including Rod Brooks, Cynthia Breazeal, Matthew Marjanovic, Paul Fitzpatrick, Bryan Adams and Aaron Edsinger.

REFERENCES

[1] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?" *Behavioral and Brain Sciences*, vol. 4, pp. 515–526, 1978.
[2] S. Baron-Cohen, *Mindblindness*. MIT Press, 1995.
[3] A. M. Leslie, "ToMM, ToBY, and Agency: Core architecture and domain specificity," in *Mapping the Mind: Domain specificity in cognition and culture*, L. A. Hirschfeld and S. A. Gelman, Eds. Cambridge University Press, 1994, pp. 119–148.
[4] D. C. Dennett, *The Intentional Stance*. MIT Press, 1987.
[5] D. Premack, ""Does the chimpanzee have a theory of mind?" revisited," in *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans.*, R. Byrne and A. Whiten, Eds. Oxford University Press, 1988.
[6] D. J. Povinelli and T. M. Preuss, "Theory of mind: evolutionary history of a cognitive specialization," *Trends in Neuroscience*, vol. 18, no. 9, 1995.
[7] D. L. Cheney and R. M. Seyfarth, "Reading minds or reading behavior? Tests for a theory of mind in monkeys," in *Natural Theories of Mind*, A. Whiten, Ed. Blackwell, 1991.
[8] J. Fodor, "A theory of the child's theory of mind," *Cognition*, vol. 44, pp. 283–296, 1992.
[9] H. Wimmer and J. Perner, "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception," *Cognition*, vol. 13, pp. 103–128, 1983.
[10] C. D. Frith and U. Frith, "Interacting minds – a biological basis," *Science*, vol. 286, pp. 1692–1695, 26 November 1999.
[11] S. Carey, "Sources of conceptual change," in *Conceptual Development: Piaget's Legacy*, E. K. Scholnick, K. Nelson, S. A. Gelman, and P. H. Miller, Eds. Lawrence Erlbaum Associates, 1999, pp. 293–326.
[12] R. Gelman, "First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples," *Cognitive Science*, vol. 14, pp. 79–106, 1990.
[13] J. Perner and B. Lang, "Development of theory of mind and executive control," *Trends in Cognitive Sciences*, vol. 3, no. 9, September 1999.
[14] A. Karmiloff-Smith, E. Klima, U. Bellugi, J. Grant, and S. Baron-Cohen, "Is there a social module? Language, face processing, and theory of mind in individuals with Williams Syndrome," *Journal of Cognitive Neuroscience*, vol. 7:2, pp. 196–208, 1995.
[15] P. Mundy and M. Sigman, "The theoretical implications of joint attention deficits in autism," *Development and Psychopathology*, vol. 1, pp. 173–183, 1989.
[16] A. Michotte, *The perception of causality*. Andover, MA: Methuen, 1962.
[17] B. Adams, C. Breazeal, R. Brooks, and B. Scassellati, "Humanoid robotics: A new kind of tool," *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 25–31, July/August 2000.
[18] B. Scassellati, "Theory of mind for a humanoid robot," *Autonomous Robots*, vol. 12, no. 1, pp. 13–24, 2002.
[19] ——, "Foundations for a theory of mind for a humanoid robot," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.
[20] G. Butterworth, "The ontogeny and phylogeny of joint visual attention," in *Natural Theories of Mind*, A. Whiten, Ed. Blackwell, 1991.
[21] J. M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bulletin & Review*, vol. 1, no. 2, pp. 202–238, 1994.
[22] G. Butterworth, "Joint attention is based on the facts of embodiment and not on a theory of mind," http://www.warwick.ac.uk/fac/soc/Philosophy/consciousness/abstracts/Butterworth.html, 2000.
[23] B. Webb, "Can robots make good models of biological behaviour?" *Behavioral and Brain Sciences*, vol. 24, no. 6, 2001.