Theory of Mind for a Humanoid Robot

Brian Scassellati

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: One of the fundamental social skills for humans is a theory of other minds. This set of skills allows us to attribute beliefs, goals, and desires to other individuals. To take part in normal human social dynamics, a robot must not only know about the properties of objects, but also the properties of animate agents in the world. This research project attempts to implement basic social skills on a humanoid robot using models of social development in both normal and autistic children.

Motivation: Human social dynamics rely upon the ability to correctly attribute beliefs, goals, and percepts to other people. This set of metarepresentational abilities, which have been collectively called a "theory of mind", allows us to understand the actions and expressions of others within an intentional or goal-directed framework. The recognition that other individuals have knowledge, perceptions, and intentions that differ from our own is a critical step in a child's development and is believed to be instrumental in self-recognition, grounding in linguistic acquisition, and possibly in the development of imaginative and creative play [4]. These abilities are also central to what defines human interactions. Normal social interactions depend upon the recognition of other points of view, the understanding of other mental states, and the recognition of complex non-verbal signals of attention and emotional state.

Previous Work: Two popular models have been proposed to explain the developmental structure of theory of mind skills in children: one from Leslie [3] and one from Baron-Cohen [1]. I have previously proposed a merger of these two models [5] that is amenable to implementation on our humanoid robot Cog (Figure 1). A variety of simple visual feature detectors (including motion, color saliency, skin color, face templates, and others) has already been implemented to provide the basic pre-attentional visual system. A context-dependent attention system combines these low-level features with high-level motivational influences [2].

Approach: The proposed structure for building theory of mind skills for our humanoid robot is a combination of the first module of Leslie's model, which is the "Theory of Body" (or ToBY) module. ToBY applies a set of naive physical laws to objects in order to distinguish animate from inanimate objects. Objects that obey purely physical laws are seen to be inanimate, while those that break the simple physical laws and display self-propelled motion are seen as animate. Animate stimuli are further processed by modules from Baron-Cohen's model. The "Eye Direction Detector" (EDD) will determine the angle of gaze of an individual and extrapolate to find the object of interest. The "Intentionality Detector" (ID) observes motion patterns and attributes the simple intentional states of desire and fear to animate objects. The "Shared Attention Mechanism" (SAM) produces representations of attentional states and allows the robot to observe objects that are under consideration by the observer as highly salient.

Difficulty: The primary difficulty in this work is integrating techniques from a wide variety of disciplines including machine vision (face recognition, gesture identification, etc.), visual-motor coordination, and machine learning.

Impact: Building a developmental model of these basic social skills on our robot may give insight into the models of Baron-Cohen and Leslie. Implementing a model of joint attention will allow us to test and evaluate that model, uncovering inconsistencies and filling out unspecified details. With an implementation of joint attention, our robot can serve as a type of experimental apparatus that has been unavailable to the cognitive science community. Effects of different environments, different experiences, or even abnormalities of development can be investigated in isolation by simply changing the parameters of our robotic platform.

These mechanisms will also allow our robot to learn from people in a natural, unconstrained manner. Just as a child learns social skills and conventions through interactions with its parents, our robot will learn to interact with people using natural, social communication.

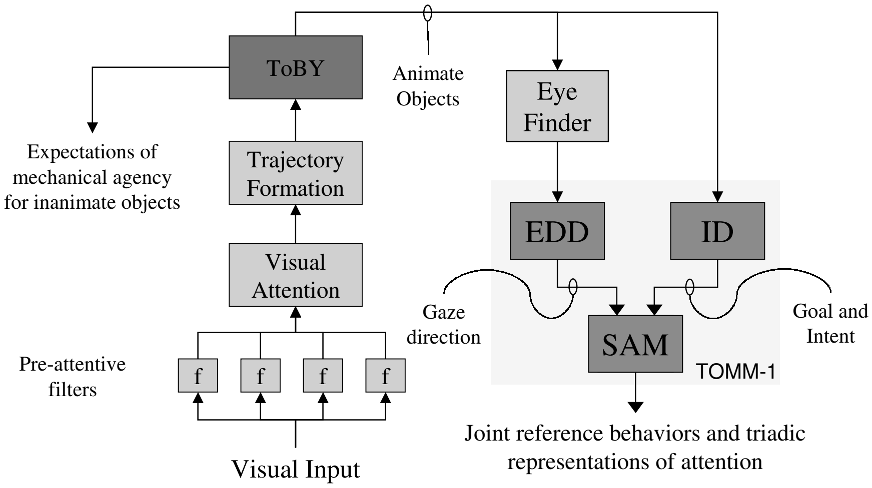

Figure 1: Proposed merger of Baron-Cohen's model and Leslie's model of the development of theory of mind skills. See text for details.

Future Work: This research is just beginning. We currently have built perceptual systems for detecting faces, eyes, visual motion, and areas of high color saliency. We also have constructed motor control systems for a variety of eye movements (saccades, smooth pursuit, and VOR/OKN), neck movements (including orienting to salient objects with efference copy), and some preliminary arm motions (such as pointing to a visual target). We have also constructed an attentional system that integrates perceptual saliency with the robot's motivations and goals to direct limited computational and motor resources. Our next step is the detection of mutual gaze and finding gaze direction.

Research Support: 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."

References:

[1] S. Baron-Cohen. *Mindblindness* MIT Press, 1995.

[2] C. Breazeal and B. Scassellati. A Context-Dependent Attention System for a Social Robot. IJCAI 99.

[3] A. Leslie. The Perception of Causality in Infants. *Perception*, 11:173–186, 1982.

[4] D. J. Povinelli and T. M. Preuss. Theory of mind: evolutionary history of a cognitive specialization. *Trends in Neuroscience*, 18:9:418–424, 1995.

[5] B. Scassellati. Theory of Mind for a Humanoid Robot. First IEEE/RSJ International Conference on Humanoid Robotics, September, 2000.