Animacy Perception and Mind Attribution in a Cognitive Architecture for Human-Robot Interaction

Samuel Spaulding
MIT Media Lab
20 Ames St. (E15-468)
Cambridge, MA 02139
samuelsp@media.mit.edu

Cynthia Breazeal
MIT Media Lab
20 Ames St. (E15-468)
Cambridge, MA 02139
cynthiab@media.mit.edu

ABSTRACT

For a cognitive architecture to be useful for Human-Robot Interaction, it will need to have a highly developed understanding of humans, their actions, and their interactions with the environment. Chief among abilities we wish to build around is the ability to understand, infer, and reason about the beliefs of others. Already work has progressed on developing such capabilities, but current work has not yet addressed its underlying origins - the ability to determine which things *have* minds and beliefs to be reasoned about. This paper outlines why it is important to develop the fundamental cognitive precursors to full theory of mind and proposes that integrating perception of animacy and mind attribution into a cognitive architecture for human-robot interaction could both provide progress towards a complete theory of mind and a path for integrating and consolidating perceptual data into symbolic data for reasoning.

Keywords

Theory of Mind, Human-Robot Interaction, Cognitive Architecture

1. INTRODUCTION

What things have minds? While philosophers continue to debate the question, humans make decisions of this sort every day. Humans attribute minds, beliefs, and intentions to other humans, animals, and even machines and computer systems (as in “My car loves me, it always starts when I need it to.” or “My phone hates it when I play that game, it always gets so slow”)

These sorts of statements are no mere metaphor. Across a wide scope of scenarios ranging from high-level, active interactions with a computer system [9] to low-level, passive perceptions of shapes on a screen [5], the tendency to adopt the intentional stance - to interpret the actions of others as a result of intentions, beliefs, and desires - is immediate, irresistible, and robust. [3]

From a cognitive science perspective, any cognitive architecture that seeks to model the human mind will have to account for this remarkably universal facet of human cognition. From the HRI perspective, the ability to conceive of agents as mentalistic is indispensible for predicting, understanding, and executing human or human-like behavior. Simply put, there is a need to consider how animacy perception and mind attribution will figure into a cognitive architecture for HRI.

2. ANIMACY PERCEPTION AND MIND ATTRIBUTION

Theory of mind is a rich and multi-faceted aspect of human social cognition that, among other things, allows us to understand and predict the mental states of others and engage in perspective taking to simulate beliefs and stimulate empathy. Baron-Cohen and others hypothesize a “theory-of-mind module” (TOMM) [1] that is responsible for attribution and analysis of others’ minds, and the HRI community has worked to extend particular aspects of theory of mind to robots in specific contexts [10, 2, 6].

Several developmental precursors to a complete theory of mind have been identified, most importantly the ability to conceive of others as mentalistic agents [7]. Underlying this *cognitive* ability is the *perceptual* ability to recognize animate movement, thus we consider animacy perception and mind attribution to be seperate, but closely linked abilities.

Several interesting qualities of animacy perception have been identified by psychologists including its remarkable resilience, immediacy, and universality [11]. Children as young as 12 months can recognize and distinguish animate movement (e.g., searching) from inanimate movement (e.g., rolling along a trajectory). [8]

Detecting animacy is an important, but by no means exclusive, component of attributing mind. To say that humans attribute minds to shapes and computers need not imply delusion. Rather than implying that “mind attribution” means to genuinely believe that something is conscious and thinking, we take a broader view and take mind attribution to specifically include *behaving as if* the system in question had a mind. To clarify - when questioned, people understand that these agentic systems do not *actually* have minds. Still, their behavior towards these systems changes drastically, depending on whether or not the system in question is perceived as an agent. [12, 13]

From this perspective, mind attribution is much more than just determining what is alive or not. It relies on a complex understanding of actions, movement, and intention.

Human-Robot Interaction ’14 Bielefeld, Germany

Such an understanding is critical for the development of theory of mind and for any system that seeks to understand and interact with humans.

3. CONNECTING PERCEPTION AND COGNITION IN COGNITIVE ARCHITECTURES

Significant efforts have been made to integrate a full-blown TOMM at the architectural level [14] and to develop stand alone computational models of detecting and classifying agentic behavior [4]. To date, however, there has been no work specifically focused on how to connect the two by integrating the perception of animacy and the attribution of other minds into a larger architecture.

Synthesizing and integrating computational models for animacy perception into a cognitive architecture, besides providing valuable social competencies, may also provide more general benefits from a design perspective. Some of the qualities we desire in a cognitive architecture are capability, generality, and flexibility. An important signal of these qualities are the higher-order capabilities that emerge from a handful of lower-level modules. We would like to have a cognitive architecture that uses relatively few directly implemented computational models, yet can give rise to complex integrative cognitive abilities. The question of how to generate the complex phenomenon of mind attribution from comparatively simple perceptual processes like animacy perception is a motivating example for the more general consideration of how to structure a cognitive architecture to maximally promote these conditions.

Furthermore, developing these competencies in the context of a larger cognitive architecture may lead to design progress on many other important cognitive capabilities that require bridging the gap from perception to cognition. Animacy perception and mind attribution is one example of two closely linked skills that fall on opposite sides of this divide, but this distinction also underlies many other cognitive abilities such as the visual grounding of symbols and tactile feedback.

Theory of mind is a highly complex ability that requires integrating many disparate types of information together. Developing and implementing the capability to recognize animate movement, model moving agents, and apply reasoning to attribute mind to them are fundamental cognitive precursors to any theory of mind. In addition, considering animacy perception and mind attribution as a motivating example could help to effectively organize a cognitive architecture for HRI.

4. CONCLUDING REMARKS

One approach to developing HRI systems relies on developing computational models for specific domains or cognitive abilities and integrating those models, often developed independently, together into a cohesive whole. A development strategy that instead begins with a general cognitive architecture and moves to adapt or extend it to specific interaction scenarios may therefore prove more fruitful in the development of a robot with the wide range of capabilities that social interaction requires. However, this approach places a higher burden on the initial design of the system. In this paper, we have highlighted why research into implementing the perception of animacy and its connection to mind attribution could both directly and indirectly benefit cognitive architectures for Human-Robot Interaction. Further exploration of this topic has the potential to provide a lower-level basis for improved theory of mind, while an implementation that connects the perception of animacy to the cognitive attribution of mind could serve as a design template for other components of the architecture that require tight integration of perception and cognition.

5. REFERENCES

[1] S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind.* MIT press, 1997.

[2] C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. L. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54(5):385–393, 2006.

[3] D. C. Dennett. *The Intentional Stance (Bradford Books).* MIT Press, Cambridge/Mass, 1987.

[4] K. Gold and B. Scassellati. A bayesian robot that distinguishes “self” from “other”. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, 2007.

[5] F. Heider and M. Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.

[6] L. M. Hiatt and J. G. Trafton. A cognitive model of theory of mind. In *Proceedings of the 10th International Conference on Cognitive Modeling*, pages 91–96, 2010.

[7] S. C. Johnson. The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, 4(1):22–28, 2000.

[8] S. C. Johnson, S.-J. Ok, et al. Actors and actions: The role of agent behavior in infants' attribution of goals. *Cognitive Development*, 22(3):310–322, 2007.

[9] C. Nass and B. Reeves. *The Media Equation: How People Treat Computers, Televisions, and New Media as Real People and Places.* Cambridge University Press, 1996.

[10] B. Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.

[11] B. J. Scholl and P. D. Tremoulet. Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8):299 – 309, 2000.

[12] E. Short, J. Hart, M. Vu, and B. Scassellati. No fair!! an interaction with a cheating robot. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 219–226. IEEE, 2010.

[13] L. Takayama. Perspectives on agency interacting with and through personal robots. In *Human-Computer Interaction: The Agency Perspective*, pages 195–214. Springer, 2012.

[14] J. G. Trafton, L. M. Hiatt, A. M. Harrison, F. Tamborello, S. S. Khemlani, and A. C. Schultz. Act-r/e: An embodied cognitive architecture for human robot interaction. *Journal of Human-Robot Interaction*, 2:30–55, 01/2013 2013.