From Psychological Intention Recognition Theories to Adaptive Theory of Mind for Robots: Computational Models

Francesca Bianco†
School of Computer Science and Electronic Engineering
& Department of Psychology
University of Essex
United Kingdom
fb18599@essex.ac.uk

Dimitri Ognibene
School of Computer Science and Electronic Engineering
University of Essex
United Kingdom
Dimitri.ognibene@essex.ac.uk

ABSTRACT

Progress in robots' application to everyday scenarios has increased the interest in human-robot interaction (HRI) research. However, robots' limited social skills are associated with decreased humans' positive attitude during HRI. Here, we put forward the idea of developing adaptive Theory of Mind (ToM) model-based systems for social robotics, able to deal with new situations and interact with different users in new tasks. Therefore, we grouped current research from developmental psychology debating the computational processes underlying ToM for HRI strategy development. Defining a model describing adaptive ToM processes may in fact aid the development of adaptive robotic architectures for more flexible and successful HRI. Finally, we hope with this report to both further promote the cross-talk between the fields of developmental psychology and robotics and inspire future investigations in this direction.

KEYWORDS: Theory of Mind; Intention Recognition; Belief Recognition; Action recognition; Prediction; User state

ACM Reference format:
Francesca Bianco and Dimitri Ognibene. 2020. From Psychological Intention Recognition Theories to Adaptive Theory of Mind for Robots: Computational Models. In *HRI'20 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 23-26, 2020, Cambridge, UK.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3371382.3378364

1 Introduction

Together with the recent introduction of robots in increasing everyday scenarios, human-robot interaction (HRI) research has rapidly evolved and various types and forms of interactions are being investigated [1-3]. However, although research in robots' industrial application has been highly reported, that in areas requiring additional social interaction is still at initial stages [4]. Indeed, the growing interest of HRI has mainly focused on the tuning of specific parameters to increase humans' positive response to robots.

ACM ISBN 978-1-4503-7057-8/20/03. https://doi.org/10.1145/3371382.3378364

An example is provided by [5] which reports an influence of the environment conditions on humans' perception of a "dancing robot" performance. Most of these systems, however, do not rely on model-based frameworks and overlook the way in which humans interact, preventing their adaptation to general situations. In fact, building robots as social agents, thus able to autonomously infer humans' intentions and beliefs and to adapt their behaviour both in various environments and when interacting with different people, remains one of the biggest challenges in robotics [6]. Previously [7, 8], we highlighted the need for a cross-talk between robotics and developmental psychology as a means to advance robots' social skills and improve HRI. Specifically, we suggested the integration of a human-inspired adaptive Theory of Mind (ToM) in the architectures for social robots to do so. In addition, four main functional advantages of equipping robots with an adaptive ToM with respect to current robotic architectures were identified, namely belief understanding and tracking, proactivity, active perception and learning (see [7] for further details). However, human ToM is characterised differently in the literature and it is yet to be fully defined [8]. To advance this debate, we thus identified a few important questions to be answered which would aid the integration of an adaptive ToM model in social robot architectures. Among these, is the understanding of which computational processes may underlie human ToM development. In the next section, a brief description of the main accounts suggested as precursors of ToM will be provided, together with tables summarising current psychology experiments addressing this topic. Finally, their importance for HRI will be highlighted to further support this cross-talk between different research fields[9].

2 Computational Models Underlying ToM

In psychology, three main accounts are utilised to describe the human ability to understand intentions from observed actions, i.e. the association, simulation and teleological theories. The action-effect association theory is the simplest one, which states that goals are inferred by "simple associations between an observed action and the effects that one's actions have produced" [10]. The simulation theory is the next more complex one, according to which "actions are understood when the observer directly matches, or mirrors, the observed action onto their own motor system" [11]. The last teleological theory states that "the outcome of an action is seen as the goal, depending on whether it is judged to justify the action in the given situation according to the rationality principle" [10]. Here, we provide a summary table (see Table 1 below) of state-of-the-art experiments from the world of developmental psychology researching infants' ability to use the different accounts during development. We hope to also further promote a successful cross-talk between the fields of HRI and developmental psychology [9, 12, 13]. Indeed, current robots' architectures generally rely on association and simulation principles to enable robots to learn about the social world, enabling them to recognize and predict actions from observing other agents performing such actions [12, 13].

Table 1. Summary table of developmental psychology experiments describing infants' cognitive ability to predict goal-directed actions supported by different computational theories

Study	Task	Findings	Supported Computational Theory
Woodward, 1998	Prediction of grasping action	After seeing a hand repeatedly grasping one of two objects, infants anticipate that the same object would be grasped again, even when the spatial location of the objects are rearranged	Association
Monroy et al., 2017	Prediction of action sequence	Observing actions, but not visual events, influenced toddlers' action choices when associated with an effect.	Association
Southgate et al., 2009	Prediction of grasping action	Infants display overlapping neural activity during execution and observation of actions, but this activation, rather than being directly induced by the visual input, is driven by infants' understanding of a forthcoming action.	Simulation
Skerry et al., 2013	Prediction of grasping action	Infants apply a general assumption of efficient action as soon as they have sufficient information (possibly derived from their own action experience) to identify an agent's goal in a given instance	Possibly Simulation for Teleological
Csibra et al., 2003	Goal attribution to uncompleted action	Infants use the principle of rational action for the interpretation and prediction of goal-directed actions, but also for making productive inferences about unseen aspects of their context	Teleological
Southgate et al., 2008	Prediction of goal-directed action	Infants appear to extend goal attribution even to biomechanically impossible actions as long as they are physically efficient	Teleological

Although several studies provided computational models of ToM based on the accounts above described [19, 20], we previously debated that such computational processes do not fully explain the way in which we are also able to infer abstract mental states (ToM components), such as intentions, desires and beliefs (see [7, 8] for more details on current robotic architectures for ToM). Similarly, we suggest that humans are able to access more complex concepts than those described by the previous accounts, which cannot be inferred through action observation. Likewise, robots' recognition of complex mental states and their understanding of the humans they interact with is limited when based on the mentioned theories. For example, although robots can recognize the action of displacing objects in a room, they do not understand the deeper mental state of the observed agent, e.g., desire to search for glasses. However, we proposed that these models may be important for and precursors of ToM. Different researchers have attempted to link one of these theories and ToM, contributing to the debate. Here, we illustrate some of these investigations in Table 2 below to both group such proposals, to the best of our knowledge, for the first time and inspire future studies in this direction. Specifically, we believe that defining whether such computational processes can cooperate to build a ToM or whether they are mutually exclusive is essential for characterising ToM and provide a truthful description of the underlying computational processes. This will in turn prove useful for the development of adaptive architectures for social robots.

Table 2. Summary of psychology experiments describing precursor computational models of human Theory of Mind ability

Study	Precursor of Theory of Mind	Proposal	Experiments
Keysers & Gazzola, 2006	Simulation as precursor of Theory of Mind	Based on neural evidence: brain areas associated with both accounts represent simulation (even though at different levels)	N/A
Gergely & Csibra, 2003	Teleological as precursor of Theory of Mind	Continuum between teleological constructs (i.e., action, goal-state and situational constraints) and mentalistic ones (i.e., intentions, desires and beliefs), with the latter supposing same computations and constructs of the former but representing more sophisticated, abstract constructs	12-month-old infants had to infer a goal state to rationalize the incomplete action, whose end state was occluded from them, as an efficient 'chasing' action
Baker et al., 2017	Teleological as precursor of Theory of Mind	Bayesian computational model for ToM based on the teleological principle: adults were suggested to follow this model to infer the mental states behind an agent's behavior using priors	'Food-trucks' scenario, using animated two-dimensional displays of an agent navigating through simple grid-worlds. Observers had to infer agents' mental states according to their trajectory
Hamlin Et al., 2013	Teleological as precursor of Theory of Mind	Bayesian computational model for ToM based on the teleological principle: infants were suggested to follow this model to infer the mental states behind an agent's behavior	Evaluation of social actions between agents (helping or harming intentions of agents during interaction)

3 Conclusion

In this paper, we grouped some key findings in developmental psychology concerning current computational theories describing intention understanding from observed actions, which have also inspired the development of architectures for social robots. In addition, we summarised some current attempts in psychology to describe the human higher-level ToM ability for mental states inference and propose both psychology and robotics studies to continue in this direction. Specifically, we would like to further stress the importance of the cross-talk between the two research fields. Indeed, understanding how infants develop such social skills may be a resource for more adaptive and complex robotic systems, improving HRI and increasing the application of robots in additional social scenarios.

REFERENCES

[1] A. Abubshait, and E. Wiese, "You Look Human, But Act Like a Machine: Agent Appearance and Behavior Modulate Different Aspects of Human-Robot Interaction," Front Psychol, vol. 8, pp. 1393, 2017.

[2] L. Sinnema, and M. Alimardani, "The Attitude of Elderly and Young Adults towards a Humanoid Robot as a Facilitator for Social Interaction. ," 2019.

[3] Alessia Vignolo, Alessandra Sciutti, Francesco Rea, and J. Michael, "Spatiotemporal Coordination Supports a Sense of Commitment in Human-Robot Interaction," 2019.

[4] T. B. Sheridan, "Human–Robot Interaction: Status and Challenges," Human Factors, vol. 58, no. 4, pp. 525-532, 2016.

[5] Cecilia Roselli, Francesca Ciardo, and A. Wykowska, "Robots Improve Judgments on Self-generated Actions: An Intentional Binding Study."

[6] J. B. G.Z. Yang, P. E. Dupont, P. Fisher, L. Floridi, R. Full, et al, "The grand challenges of science robotics," Science Robotics, vol. 3, pp. eaar7650, 2018.

[7] F. Bianco, Ognibene, D., "Functional Advantages of an adaptive Theory of Mind for robotics: a review of current architectures."

[8] F. Bianco, Ognibene, D., "Transferring adaptive Theory of Mind to social robots: insights from developmental psychology."

[9] S. Lemaignan, and P. Dillenbourg, "Mutual Modelling in Robotics: Inspirations for the Next Steps," ACM/IEEE International Conference on Human-Robot Interaction, vol. 2015, pp. 303-310, 03/02, 2015.

[10] G. G. Csibra G, "'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans," acta psychol, vol. 124, no. 1, pp. 60-78, 2007.

[11] V. Southgate, M. H. Johnson, and G. Csibra, "Infants attribute goals even to biomechanically impossible actions," Cognition, vol. 107, no. 3, pp. 1059-69, Jun, 2008.

[12] A. Cangelosi, and M. Schlesinger, "From Babies to Robots: The Contribution of Developmental Robotics to Developmental Psychology," Child Development Perspectives, vol. 12, no. 3, pp. 183-188, 2018.

[13] Y. Demiris, and A. Dearden, "From motor babbling to hierarchical learning by imitation: a robot developmental pathway," Proceedings of the 5th International Workshop on Epigenetic Robotics, pp. 31-37, 2005.

[14] A. L. Woodward, "Infants selectively encode the goal object of an actor's reach," Cognition, vol. 69, no. 1, pp. 1-34, Nov, 1998.

[15] C. Monroy, S. Gerson, and S. Hunnius, "Infants' Motor Proficiency and Statistical Learning for Actions," Front Psychol, vol. 8, pp. 2174, 2017.

[16] V. Southgate, M. H. Johnson, T. Osborne, and G. Csibra, "Predictive motor activation during action observation in human infants," Biol Lett, vol. 5, no. 6, pp. 769-72, Dec 23, 2009.

[17] A. E. Skerry, S. E. Carey, and E. S. Spelke, "First-person action experience reveals sensitivity to action efficiency in prereaching infants," Proceedings of the National Academy of Sciences, vol. 110, no. 46, pp. 18728-18733, 2013.

[18] G. Gergely, and G. Csibra, "Teleological reasoning in infancy: the nai;ve theory of rational action," Trends Cogn Sci, vol. 7, no. 7, pp. 287-292, Jul, 2003.

[19] L. Hiatt, A. Harrison, and J. Trafton, Accommodating Human Variability in Human-Robot Teams through Theory of Mind, 2011.

[20] A. F. T. Winfield, "Experiments in Artificial Theory of Mind: From Safety to Story-Telling," Frontiers in Robotics and AI, vol. 5, no. 75, 2018-June-26, 2018.

[21] C. Keysers, and V. Gazzola, "Integrating simulation and theory of mind: From self to social cognition," Trends in Cognitive Sciences, vol. 11, no. 5, pp. 194-196, 2007.

[22] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," Nature Human Behaviour, vol. 1, no. 4, pp. 0064, 2017.

[23] J. Hamlin, T. Ullman, J. Tenenbaum, N. Goodman, and C. Baker, "The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model," Dev Sci, vol. 16, no. 2, pp. 209-226, Mar, 2013.