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:

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.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author(s).

HRI ’20 Companion, March 23–26, 2020, Cambridge, United Kingdom
© 2020 Copyright is held by the owner/author(s).

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, proactiveness, 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

<table>
<thead>
<tr>
<th>Study</th>
<th>Task</th>
<th>Findings</th>
<th>Supported Computational Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodward, 1998</td>
<td>Prediction of grasping action</td>
<td>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 re arranged</td>
<td>Association</td>
</tr>
<tr>
<td>Mervis et al., 2017</td>
<td>Prediction of action sequence</td>
<td>Observing actions, but not visual events, influenced toddlers’ action choices when associated with an effort</td>
<td>Association</td>
</tr>
<tr>
<td>Southgate et al., 2009</td>
<td>Prediction of grasping action</td>
<td>Infants display overlapping neural activity during execution and observation of actions, but this activation, rather than being directly induced by the visual agent, is driven by infants’ understanding of a forthcoming action.</td>
<td>Simulation</td>
</tr>
<tr>
<td>Sherry et al., 2013</td>
<td>Prediction of grasping action</td>
<td>Infants apply a general assumption of efficient actions as soon as they have sufficient information (possibly derived from their own action experience) to identify an agent’s goal in a given instance</td>
<td>Possibility Simulation for Teleological</td>
</tr>
<tr>
<td>Calia et al., 2005</td>
<td>Goal attribution to completed action</td>
<td>Infants use the principle of rational actions for the interpretation and prediction of goal-directed actions, but also for making predictive inferences about unobserved aspects of their context</td>
<td>Teleological</td>
</tr>
<tr>
<td>Southgate et al., 2008</td>
<td>Prediction of goal-directed action</td>
<td>Infants appear to extend goal attribution even to biomechanically impossible actions as long as they are physically efficient</td>
<td>Teleological</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Precursor of Theory of Mind</th>
<th>Proposal</th>
<th>Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keysers &amp; Oszust, 2006</td>
<td>Simulation as precursor of Theory of Mind</td>
<td>Based on neural evidence: brain areas associated with both accounts represent simulation (even though at different levels)</td>
<td>N/A</td>
</tr>
<tr>
<td>Oergel &amp; Caba, 2003</td>
<td>Teleological as precursor of Theory of Mind</td>
<td>Continuum between teleological constructs (i.e., action, goal-state and situational constructs) and simulations: ones (i.e., intentions, desires and beliefs), with the latter opposing some computations and constructs of the former but representing more sophisticated, abstract constructs</td>
<td>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</td>
</tr>
<tr>
<td>Baker et al., 2017</td>
<td>Teleological as precursor of Theory of Mind</td>
<td>Bayesian computational model for ToM based on the teleological principle: infants were suggested to follow this model to infer the mental state behind an agent’s behavior using priors</td>
<td>‘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</td>
</tr>
<tr>
<td>Hamlin et al., 2013</td>
<td>Teleological as precursor of Theory of Mind</td>
<td>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</td>
<td>Evaluations of social actions between agents (helping or harming intentions of agents during interaction)</td>
</tr>
</tbody>
</table>

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.

137
REFERENCES