AToM: An Analogical Theory of Mind

Irina Rabkina
Qualitative Reasoning Group, Northwestern University
Evanston, IL 60208
irabkina@u.northwestern.edu

Abstract

Theory of Mind (ToM) has been well studied in psychology. It is what gives adults the ability to predict other people's beliefs, desires, and related actions. When ToM is not yet developed, as in young children, social interaction is difficult. A cognitive system that interacts with humans on a regular basis would benefit from having a ToM. In this extended abstract, I propose a computational model of ToM, Analogical Theory of Mind (AToM), based on Bach's [2012, 2014] theoretical Structure-Mapping model of ToM. Completed work demonstrates the plausibility of AToM. Future steps include a full implementation and test of AToM.

1 Introduction

Humans are inherently social creatures. In fact, it has been suggested that our need for social interaction is responsible for our large brains and incredible language abilities [e.g. Reader and Laland, 2002]. If artificial intelligence systems are to be integrated into our society, then they, too, must have the social capabilities available to us.

Theory of Mind (ToM) is one example of a capability necessary for social interaction. ToM, sometimes referred to as mind reading, is the ability to predict others' desires, beliefs, and other mental states even when they may be different from our own. While some evidence of ToM exists in a other highly social animals, such as dolphins and apes [e.g. Krupenye *et al.* 2016], the extent to which we use and rely on ToM seems to be uniquely human.

The philosopher Theodore Bach [2011, 2014] proposed one theory, based in Structure-Mapping Theory [SMT, Gentner, 1983], of how ToM is developed and used by humans. This extended abstract describes a computational cognitive model of ToM, Analogical Theory of Mind (AToM), which is based on Bach's theory. Previous work, which shows how processes which play a role in ToM development can be used to train AToM, is presented. Finally, future directions are discussed.

2 Analogical Theory of Mind (AToM)

AToM is based on the Structure-Mapping Theory of ToM proposed by Bach [2011, 2014]. It is built on top of the Structure-Mapping Engine [SME, Forbus *et al.* 2016], a computational model of SMT; the SAGE model of analogical generalization [McLure *et al.* 2010]; and the MAC/FAC model of analogical retrieval [Forbus *et al.* 1995]. AToM assumes a long term memory (LTM) of predicate calculus cases that can be retrieved via MAC/FAC. These cases represent memories of life experiences.

When a case which requires ToM reasoning is encountered, AToM retrieves a relevant case from LTM using MAC/FAC. If the retrieved case is a generalized schema, it is applied as a rule. If the retrieved case is a single event, an interim generalization is created in working memory [Kandaswamy *et al*. 2014]. While standard interim generalizations are created via SAGE, a slightly different process is involved for AToM's generalizations. Candidate inferences from the retrieved case are projected onto the target case and, where necessary, portions of the target case are re-represented based on the candidate inferences. This interim generalization is used for ToM reasoning. AToM then asks for feedback in natural language [using EA-NLU, Tomai and Forbus, 2009]. This is analogous to a person receiving feedback on their reasoning by interacting with others. If the reasoning was correct, AToM uses SAGE to generalize the original probe with the retrieved case, and stores the new generalized case in LTM. Otherwise, it uses MAC/FAC to find a better match (again, given the feedback) and generalizes with the new match. In this way, schemas become more and more generalized, and ToM abilities continue to improve.

When complete, AToM will be integrated into the Companion cognitive architecture [Forbus *et al.* 2009]. This will enhance a Companion's interactive abilities.

3 Previous Work

Here, I describe two completed studies that show AToM is a plausible model of ToM.

3.1 Pretense

Pretense plays a role in ToM development [Weisberg, 2015], so it can be used as a simplified domain to test AToM. Specificially, we tested AToM's mechanisms for candidate inference verification and re-representation by modeling two studies of pretend play [Rabkina and Forbus, in prep].

In the model, when a pretend scenario is encountered, a schema of its real-life equivalent is retrieved from LTM. The two are compared via SME, and candidate inferences are projected from the schema to the pretend scenario. Pretend play is considered successful when the proper candidate inferences are accepted and the pretend scenario is properly transformed.

The results of the model are consistent with both successful and failed pretense in children. This provides support for the mechanisms behind AToM.

3.2 ToM Training Study

The remaining steps of AToM (i.e. retrieval, generalization, and reasoning) were tested by modeling a study in which children were able to learn some aspects of ToM in the lab [Hoyos *et al.* 2015; Rabkina *et al.* 2017]. Using only three vignettes as inputs, AToM successfully performed a series of false belief tasks, a standard test for ToM development.

This study shows not only that the mechanisms behind AToM are cognitively plausible, but also that the ToM reasoning that results is comparable to the reasoning performed by humans. This suggests that software, such as the Companion cognitive architecture, which includes AToM for ToM reasoning will be able to interact with users in a more human-like way.

4 Future Directions

A major challenge for this project is testing AToM in a way that is both computationally appropriate and cognitively plausible. In the lab, ToM is usually tested using simple tasks that serve as a proxy for stages of ToM development. The ability to complete these tasks is not sufficient to have a complete ToM. Identifying additional cognitive tasks is currently in progress, with further testing being future work.

Acknowledgments

This research was supported by the Socio-Cognitive Architectures for Adaptable Autonomous Systems Program of the Office of Naval Research, N00014-13-1-0470.

References

[Bach, 2011] Theodore Bach. Structure-mapping: Directions from simulation to theory. *Philosophical Psychology*, 24(1): 23-51. 2011.

[Bach, 2014] Theodore Bach. A Unified Account of General Learning Mechanisms and Theory-of-Mind Development. *Mind & Language*, 29(3): 351-381. 2014.

[Gentner, 1983] Dedre Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170. 1983.

[Forbus *et al*. 1995] Kenneth D. Forbus, Dedre Gentner, and Keith Law. MAC/FAC: A Model of Similarity-based Retrieval. *Cognitive Science*, 19:141-205. 1995.

[Forbus *et al.* 2009] Kenneth D. Forbus, Matthew Klenk, and Thomas Hinrichs. Companion cognitive systems: Design goals and lessons learned so far. *IEEE Intelligent Systems*, 24(4), 36-46. 2009.

[Forbus *et al.* 2016] Kenneth D. Forbus, Ronald W. Ferguson, Andrew Lovett, and Dedre Gentner. Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, DOI: 10.1111/cogs.12377, pp 1-50. 2016.

[Hoyos *et al.* 2015] Christian Hoyos, William S. Horton, and Dedre Gentner. Analogical comparison aids false belief understanding in preschoolers. In *Proceedings of the Cognitive Science Society.* 2015.

[Kandaswamy *et al.* 2014] Subu Kandaswamy, Kenneth Forbus, and Dedre Gentner. Modeling Learning via Progressive Alignment using Interim Generalizations. *Proceedings of the Cognitive Science Society.* 2014.

[Krupenye *et al.* 2016] Christopher Krupenye, Fumihiro Kano, Satoshi Hirata, Josep Call, and Michael Tomasello. Great apes anticipate that other individuals will act according to false beliefs. *Science* 354, no. 6308: 110-114. 2016.

[McLure *et al.* 2010] Matthew McLure, Scott Friedman, and Kenneth Forbus. Learning concepts from sketches via analogical generalization and near-misses. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.* Portland, OR. 2010.

[Rabkina and Forbus, in prep] Irina Rabkina and Kenneth D. Forbus. An Analogical Model of Pretense. In prep.

[Rabkina *et al*. 2017] Irina Rabkina, Clifton McFate, Kenneth D. Forbus, and Christian Hoyos. Towards a Computational Analogical Theory of Mind. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* London, England. 2017.

[Reader and Laland, 2002] Simon M. Reader and Kevin N. Laland. Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*. 99.7:4436-4441. 2002.

[Tomai and Forbus, 2009] Emmett Tomai and Kenneth Forbus. EA NLU: Practical Language Understanding for Cognitive Modeling. *Proceedings of the 22nd International Florida Artificial Intelligence Research Society Conference*. Sanibel Island, Florida. 2009.

[Weisberg, 2015] Deena S. Weisberg. Pretend play. *Wiley Interdisciplinary Reviews: Cognitive Science* 6.3: 249-261. 2015.