How do Infants Recognize and Understand others' Mental States: Recent Accounts and A Debate

Xinyi Yang
Department of Automation
Tsinghua Universtiy
xy-yang21@mails.tsinghua.edu.cn

Abstract

There has been a lot of discussion about the nature mechanism of infants' recognizing and understanding of other agents' mental states, among which two-system account and one-system account are the most and longest debated. A variety of psychological and neuroscientific experiments have been done to provide evidence for theorists' respective theories. This essay will review these accounts and the debate, also including my personal opinions, leading to a more comprehensive understanding and a clearer future direction for research of ToM in my wish.

1 Introduction

Humans are among the most social species.[9] From infancy, we constantly interact with other people, and in contrast to other species, we are preferentially drawn to their conspecifics to learn through social learning.[14] Maybe it is the social living habit that makes us also think socially. Humans have a strong inclination to understand behaviors by mental states such as beliefs, desires, emotions, and intentions, which is known as theory of mind (ToM).[20] For example, when a man follows a woman but hides when the woman turns back, there is a good chance that the man harbors evil designs. This ability plays a critical role in everyday social life and is associated with a wide range of positive interactions, such as help or cooperation.

Research on this ability in infants has led to a heated debate about the exact nature of infants' ToM understanding and psychological reasoning more generally, which is the focus of this essay. In Sec. 2, several recent accounts for infants' capacity of mindreading will be introduced, among which the one-system account and the two-system account have been developed for a long time and discussed more. Then the main divergence in neurological evidence and false-belief experiments between these two accounts will be reviewed in Sec. 3 with some of my own arguments.

2 Recent Accounts for Infants' ToM

2.1 Two-System Account

According to Low and Watts [12], by two-system account, an efficient, unconscious, and inflexible mind-reading system supports indirect looking responses; this system is shared by infants and adults and ascribes the belief-like state of registration, which is a proxy for belief [3]. In contrast, a flexible and conscious but inefficient mind-reading system supports direct, verbal false-belief predictions; this system ascribes complex mental states to people and is emergent from age 4 as language, executive functions, and meta-representational skills develop [2, 19]. The processes that drive the early-developing system may be direct cues like line of sight, so that rapid online mindreading may occur with limited resources to process.[13] For example, if a man puts a ball in one location. And then in his absence, the ball is moved to another location, the early-developing system can predict, by considering where the man last encountered and registered the object, that upon his return he will look for it in its original location. In other words, the early-developing system can handle simple

false-belief understanding by tracking what information is available to agents about objects' locations and properties. However, the flexible system does not depend on the immediate availability of cues about what a target witnessed, but may place some demands on working memory, attention, and executive function.[6]

2.2 One-System Account

By one-system account, there is only a single system underlies human psychological reasoning, which emerges early in infancy and is mentalistic in nature. Infants start with a set of conceptual primitives, *e.g.*, *thinks*, *likes*, *is aware of*, and maybe *tells*, together with some simple rules for determining the application of these concepts.[7] For example, if an infant sees a ball placed into a box in the presence of a man, the infant will encode the man *be aware of* the ball is inside the box, and infer that the man *thinks* the ball is in the box. And this system is constrained by a principle of rationality, *i.e.*, *consistency* that agents act in a manner consistent with their mental states, and *efficient* that agents act with as little effort as possible.[4] Moreover, one-system theorists assume that infants' psychological reasoning is qualitatively similar to that of older children and adults, so that some failures in complex causal structures are caused by limited executive functioning and pragmatic understanding.

2.3 Deflationary Account

Completely different from the two above, deflationary account assumes that infants' psychological reasoning can be explained in non-mentalistic terms with action-effect associations, low-level stimulus factors, or behavioral rules. Heyes [10] proposed a hypothesis that the infants' appreciating of false beliefs in VoE experiments is a function of the degree to which the observed and remembered or expected low-level properties of the test stimuli, *e.g.*, colors, shapes and movements, are novel with respect to events encoded by the infants earlier in the experiment. Similarly, Ruffman [16] argued that the excellent statistical learning skills and attention to human faces and motion enable infants' very good understanding of behavior.

After introducing three recent accounts briefly, we can see that many aspects of the exact mechanism under infants' understanding of ToM have not been fully explored. As there is a long and heated debate between two-system theorists and one-system theorists, which has inspired a variety of false-belief experiments settings and some studies searching for indirect evidence such as neurological systems. Reviewing what the debate focuses on may help with future research.

3 A Debate between Two-System and One-System

3.1 Verbal and Non-verbal Measures in False-Belief Experiments

False-belief tasks, often considered a litmus test of mature ToM, explicitly require infants to ascribe a subjective mis-representation to another agent and explain/predict their actions accordingly.[15] On the one hand, in traditional verbal tasks [21], for example, infants watched a target placed at location X as Maxi witnessed, and the target moved to location Y in Maxi's absence. Then infants were asked to predict where Maxi would search for the target. Results show that many 4-year-olds answered that Maxi would look in location X, indicating that they captured Maxi's false belief; but by contrast, most 3-year-olds said Maxi would look in location Y, indicating that the false belief was not recognized. On the other hand, although there is no false-belief understanding found in 3-year-olds' verbal predictions, in various non-traditional nonverbal tasks, the results changed. For example, in one VoE task [17] (Fig. 1), eighteen-month-olds first received familiarization trials in which an agent always hid a small key in the bottom piece of a two-piece penguin and then assembled it when the two-piece penguin was placed side by side with another one-piece penguin in different containers on the table. In the test trials, the experimenter assembled the two-piece penguin, placed it under a transparent cover, and then placed the one-piece penguin under a non-transparent cover when the agent was absent. Subsequently, the agent returned with the key and reached for one of the covers. Results show that when the agent reached for the non-transparent cover (the agent mistakes the two-piece penguin as it look exactly the same as the one-piece penguin in the familiarization trials), infants looked significantly longer. And the situation reversed when the agent witnessed the

experimenter's actions. It is demonstrated that infants can reason about the actions of an agent who mistakes one object for a different object.

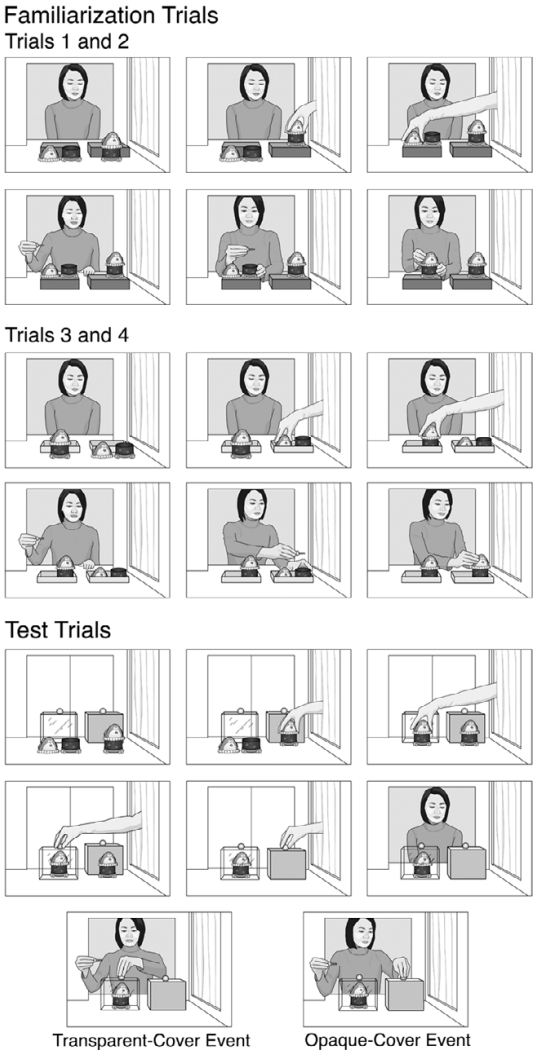

Figure 1: Familiarization and test trials shown in the false-belief condition.

In two-system theorists' opinion, the puzzle that infants are sensitive to others' false beliefs when responding in some ways while they treat false beliefs as impossible when responding to the same situation in other ways, can be solved by supposing that mindreading architecture involves at least two systems with complementary tradeoffs between efficiency and flexibility.[13] The early-developing system would support understanding mistakes in others' representations in non-traditional tasks, which are characterized as relational attitudes, such as objects and locations or other properties, so that infants could recognize the key would be placed in the wrong penguin. But traditional tasks are in need of more flexible mental representations provided by the later-developing system, where belief is characterized as a propositional attitude, that is, Maxi believes that the target is in location X. However, according to one-system theorists, the reason why infants failed in traditional tasks is that the test question was not directed at the infants, no response-generation was initiated and no prepotent incorrect response had to be inhibited.[18] When processing demands are appropriately reduced, infants can recognize false beliefs in non-traditional tasks like Fig. 1, and even correctly answer the test question themselves [1]. Furthermore, from a theoretical point of view, despite fundamental disagreements in some respects, these two kinds of accounts converge on the claim that different implicit tasks all measure the same basic capacity. But according to Dörrenberg et al. [8], it fails to find robust evidence for either replicability or convergent validity of these tasks, therefore experimental results from non-traditional tasks may need to be reconsidered.

3.2 Indirect Evidence in Neurological Systems

According to Low et al. [13], studying the temporal course of behavioral and neural activity associated with tracking belief-like states versus ascribing belief in real settings will illuminate circumstances in which information might pass between systems. But as far as I know, there is no evidence for the claim that two systems with distinct neurological substrates and computational capacities yet in the neuroscientific domain.[18] Investigations with adults using either functional near-infrared spectroscopy (*e.g.*, Hyde et al. [11]) or functional magnetic-resonance imaging (*e.g.*, Bardi et al. [5]), indicate that traditional and non-traditional tasks engage similar regions with the temporal-parital junctions (TPJ). Then similar evidence in neurological systems has been found in infants, providing no support for maked discontinuities in the development of mindreading.

However, from an intuitive point of view, when we recognize other people's mental states, there are always two processes: intuitive feeling and hindsight inference. The information gained from the later process is used to correct the feelings of the previous process. Whether there are separate modules that are each in responsible for a process in the TPJ? Whether there exists obvious information passing among them? And whether their developments occur at different stages of life? More evidence from neurological systems is needed.

4 Conclusion

This essay introduces three recent accounts for infants' ToM, and reviews the debate between two mainstream accounts, two-system account and one-system account, in two aspects, traditional and non-traditional tasks and neurological systems. Same psychological experiments results are interpreted differently by theorists, and there is no direct neuroscientific evidence provided to prove which account is totally wrong. To fully explore the exact mechanism under infants' recognizing and understanding of others' mental states, more research would be required.

References

[1] F Antilici and R Baillargeon. 2.5-year-olds pass an explicit unexpected-transfer false-belief task when processing demands are reduced. In *Biennial Meeting of the International Congress of Infant Studies*, 2020. 3

[2] Ian Apperly. *Mindreaders: the cognitive basis of" theory of mind"*. Psychology Press, 2010. 1

[3] Ian A Apperly and Stephen A Butterfill. Do humans have two systems to track beliefs and belief-like states? *Psychological review*, 116(4):953, 2009. 1

[4] Renée Baillargeon, Rose M Scott, and Lin Bian. Psychological reasoning in infancy. *Annual review of psychology*, 67:159–186, 2016. 2

[5] Lara Bardi, Charlotte Desmet, Annabel Nijhof, Jan R Wiersema, and Marcel Brass. Brain activation for spontaneous and explicit false belief tasks overlaps: new fmri evidence on belief processing and violation of expectation. *Social cognitive and affective neuroscience*, 12(3): 391–400, 2017. 4

[6] Stephen A Butterfill and Ian A Apperly. How to construct a minimal theory of mind. *Mind & Language*, 28(5):606–637, 2013. 2

[7] Peter Carruthers. Two systems for mindreading? *Review of Philosophy and Psychology*, 7: 141–162, 2016. 2

[8] Sebastian Dörrenberg, Hannes Rakoczy, and Ulf Liszkowski. How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, 46:12–30, 2018. 3

[9] Jean-François Gariépy, Karli K Watson, Emily Du, Diana L Xie, Joshua Erb, Dianna Amasino, and Michael L Platt. Social learning in humans and other animals. *Frontiers in neuroscience*, 8:58, 2014. 1

[10] Cecilia Heyes. False belief in infancy: A fresh look. *Developmental science*, 17(5):647–659, 2014. 2

[11] Daniel C Hyde, Mariana Aparicio Betancourt, and Charline E Simon. Human temporal-parietal junction spontaneously tracks others' beliefs: A functional near-infrared spectroscopy study. *Human Brain Mapping*, 36(12):4831–4846, 2015. 4

[12] Jason Low and Joseph Watts. Attributing false beliefs about object identity reveals a signature blind spot in humans efficient mind-reading system. *Psychological Science*, 24(3):305–311, 2013. 1

[13] Jason Low, Ian A Apperly, Stephen A Butterfill, and Hannes Rakoczy. Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3):184–189, 2016. 1, 3, 4

[14] Elizabeth E Price, Lara A Wood, and Andrew Whiten. Adaptive cultural transmission biases in children and nonhuman primates. *Infant Behavior and Development*, 48:45–53, 2017. 1

[15] Hannes Rakoczy. Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, 1(4):223–235, 2022. 2

[16] Ted Ruffman. To belief or not belief: Childrens theory of mind. *Developmental review*, 34(3): 265–293, 2014. 2

[17] Rose M Scott and Renée Baillargeon. Which penguin is this? attributing false beliefs about object identity at 18 months. *Child development*, 80(4):1172–1196, 2009. 2

[18] Rose M Scott, Erin Roby, and Renée Baillargeon. 11 how sophisticated is infantst theory of mind? *The cambridge handbook of cognitive development*, page 242, 2022. 3, 4

[19] Bo Wang, Jason Low, Zhang Jing, and Qu Qinghua. Chinese preschoolers implicit and explicit false-belief understanding. *British Journal of Developmental Psychology*, 30(1):123–140, 2012. 1

[20] Henry M Wellman. *Making minds: How theory of mind develops*. Oxford University Press, 2014. 1

[21] Heinz Wimmer and Josef Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128, 1983. 2