Theory of Mind

Laura Franchin
Department of Psychology and Cognitive Sciences, University of Trento, Trento, Italy

Abstract

The theory of mind (ToM) refers to how people understand their own thoughts and feelings and those of other beings. It is a crucial cognitive mechanism for social interactions and communication. It helps us to predict, to explain, and to manipulate behaviors or mental states. Moreover, this skill is shared by almost all human beings beyond early childhood.

The literature presents different explicit false-belief tasks as a means of investigating ToM in children (e.g., one of the most famous is known as the Sally-Anne task). Although children younger than 4 years usually fail in these explicit tasks, it cannot be excluded that some less complex forms of understanding mental states develop earlier. So, in order to investigate the precursors that anticipate the emergence of a more mature representational system, many recent studies on infants' beliefs have demonstrated, in the last decade, a very early sensitivity specifically to the false beliefs of others by using implicit looking-time tasks. This entry starts with the definition of the theory of mind and its history, before moving on to summarize developmental research in this area. Finally, it focuses on the relation between theory of mind and the *possible* with some reflections on how an increasing consciousness of the variety of situations that the possible presents to us could allow people to choose the best alternative for themselves and others.

Keywords

Theory of mind · ToM · Mind reading · Mentalizing · Folk psychology · Possible

Definition and History

The theory of mind (also known by the acronym ToM) refers to how people understand their thoughts and feelings and those of other beings. This ability is also known as "mind reading," "mentalizing," or "folk psychology." It is one of the most important cognitive mechanisms underpinning social interaction and communication. It helps people to predict, to explain, and to manipulate behavior or mental states. ToM implies deriving what the others think from signs, such as facial expressions, body movements, and utterances. This skill is shared by almost all human beings beyond early childhood.

Nevertheless, these mental representations are not always correct; they might turn out to be wrong which would be referred to as "false beliefs." This could happen for several reasons, such as an event having occurred without our knowledge. Thus, we can recognize that our belief was false, subsequently acquiring new information or updating our belief. This meta-representational ability allows us to realize that our beliefs can vary on the basis of the knowledge we have available (e.g., Mitchell and Lacohée 1991; Wellman 2017).

Although the philosophers have been discussing mind reading for centuries, it is only in the last 40 years that scientific research has focused on the intensive study of this topic. The term "theory of mind" appeared in scientific literature in 1978 when Premack and Woodruff published "Does the chimpanzee have a theory of mind?". The authors defined ToM as the individual attribution of "mental states to himself and others (either to conspecifics or to other species as well)" (p. 515). In the same year, also Fodor (1978) published a philosophical paper on "Propositional attitudes" where he wrote about the "representational theory of mind." The author considered the mind as "an organ whose function is the manipulation of representations and these, in turn, provide the domain of mental processes and the (immediate) objects of mental states" (p. 521). These three authors used the term "theory of mind" referring to a system of inferences, not directly observable states, which is used to make

predictions about the behavior of other beings. Soon the term appeared in developmental psychology (see Bretherton et al. 1981) where it became a common label in its field with the publication of the book *Developing Theories of Mind* in 1988, edited by Astington, Harris, and Olson. This book presented a special collection of empirical reports and conceptual analyses discussed in two important conferences which occurred 2 years before (i.e., in May 1986, International Conference on Developing Theories of Mind, University of Toronto, Toronto, and in June 1986, Workshop on Children's Early Concept of Mind, St. John's College, Oxford). Over the years this term has been used in a wide variety of ways that today is impossible to condense into a precise definition. Doherty (2009), for example, defined ToM as an "umbrella term" for children's understanding of mental states that are belief and desire.

After a definition of the theory of mind and its historical context, this entry will continue with a brief review of empirical research on the development of the theory of mind. Finally, it will consider the relation between ToM and the possible.

Research on ToM

"Do children have a theory of mind?"

A central issue in developmental psychology has been to understand how children begin to realize that they and others have mental states, which can sometimes be true – corresponding with reality – but may also be false. In addition, how and when children realize that these states of mind drive their actions and those of others (e.g., Baron-Cohen 1991; Lewis and Mitchell 1994; Moore 1996) are also fundamental. Investigating how a ToM develops could offer important insights into the study of social cognition. Several pieces of research have shown that by the end of their preschool years, at around 3–4 years, children have these skills and begin to understand that individuals might have false beliefs about reality (e.g., Astington et al. 1988; Frye and Moore 1991; Wellman et al. 2001; Whiten 1991). Furthermore, these skills are universal because they are present in various cultures all around the world (Callaghan et al. 2005), and it seems that they are uniquely human because different adaptations of false-belief tasks have documented consistent failure in chimpanzees and other great apes (Call and Tomasello 2008).

One of the first and most famous false-belief tests for investigating ToM in children was used by Wimmer and Perner (1983). The authors made up different stories, acted out with dolls and props, such as the following: Maxi is a child who puts some chocolate into cupboard x. In his absence, his mother moves the chocolate from cupboard x into another cupboard y. Participants have to point to the cupboard where Maxi will look for the chocolate when he returns. Only when children are able to represent Maxi's wrong belief, which is "chocolate is in cupboard x," despite knowing that the chocolate is in cupboard y, will they be able to correctly indicate cupboard x. Wimmer and Perner (1983) found a strong age-based trend: the children began to pass the task at around 4–5 years (at this age level, however the majority of children pointed wrongly to the actual location y, while almost all 6–9-year-old children correctly indicated location x).

From the end of the 1970s to the end of the 1980s last century, we witnessed an increase in empirical investigations focused on the study of the ability to attribute mental states to others and its development from the second year of life (e.g., Baron-Cohen et al. 1985; Bretherton et al. 1981; MacNamara et al. 1976; Shultz and Cloghesy 1981; Shultz et al. 1980). New stories for the false-belief task were created or adapted from the original story. For example, Baron-Cohen et al. (1985) created another famous false-belief test, known as the Sally-Anne task. Sally and Anne are the two main doll characters in the story. After checking that the children know which doll is Sally and which is Anne (*Naming Question*), Sally first places a marble into her basket. Then she leaves the scene. At this point, Anne takes the marble out of Sally's basket and hides it in her own basket. Sally returns and the experimenter asks the child the key *Belief Question*: "Where will Sally look for her marble?". If the children indicate the previous

location of the marble – thinking that Sally believes that the marble is in her own basket – then they pass the Belief Question. On the contrary, if they point to the marble's current location, then they fail the question by not taking into account the doll's belief. Two further control questions confirm these conclusions if children answer correctly: "Where is the marble really?" (Reality Question) and "Where was the marble in the beginning?" (Memory Question). With these questions the experimenter could be sure that the child has both knowledge of the real current location of the marble and an accurate memory of the previous location (Baron-Cohen et al. 1985).

Although children younger than 4 years old usually fail these explicit tasks, it cannot be excluded that some less complex forms of understanding mental states develop earlier. Consequently, in order to investigate the precursors that anticipate the emergence of a more mature representational system, new empirical studies on infants' beliefs have demonstrated, in the last decade, a very early peculiar sensitivity to the false beliefs of others by using implicit looking-time tasks.

If infants are presented with a typical FB task with a change of location from box x to box y and the agent returning to the scene, infants usually show anticipatory behavior searching for box x and look longer when the agent reaches box y rather than box x; moreover, in true-belief control conditions in which the agents witnessed the object's transfer, infants show a reverse pattern, that is anticipatory searching for box y and longer looking times when the agent reaches box x (e.g., Baillargeon et al. 2010).

Many recent studies, which have investigated infants' looking behavior, suggest their ability to reason about other agents' false beliefs. However, it is important to highlight that there is considerable disagreement in literature over the question whether these results show that infants have a concept of belief similar to the one assessed by explicit tasks years later, as reported by Rakoczy (2012, 2017; see also Apperly and Butterfill 2009). Heyes (2014) provided a review of more than 20 experiments on infant false beliefs (e.g., Onishi and Baillargeon 2005; Senju et al. 2011; Song and Baillargeon 2008; Southgate et al. 2007; Surian et al. 2007; Surian and Geraci 2012; Kovács et al. 2010) and argued that positive findings reported by some of these studies are due to domain-general processes, for example, retroactive interference in memory for object location. The criticisms advanced by Heyes certainly open exciting avenues for new research aimed at understanding if infants are really able to appreciate others' states of mind.

ToM and the Possible

There is a very strong relation between ToM and the possible. ToM is based on the possible. When an adult or a child reasons about others' mental states, independently of their own true or false beliefs, they are thinking about the possibility that others could carry out specific behavior or reason in a certain way. Nobody has an absolute certainty about others' thoughts and feelings and sometimes not even about their own. Therefore, the possible plays a central role. To be aware that it is possible to consider a wide range of "beliefs" could enable us to be more functional in attributing mental states to others, also increasing our own mental flexibility.

The possible might be studied from different perspectives. Theories and research on the possible have focused on four key areas – possible worlds, possible selves, possible pasts, and possible futures (for a detailed description of these areas, see Glăveanu 2018). The theory of mind might be strongly linked in particular to the area of the possible selves. This area refers to conceptions of one's self in future states. It's an experience that leads a person, as an agent, through imaginative explorations in past or future situations (e.g., Erikson 2007; Glăveanu 2018; Zittoun and de Saint-Laurent's 2015). Obviously, we can extend this reasoning also considering not only possible selves but also possible others. All humans do not have a predetermined nature; they are beings open to the possible, at different degrees. Possible selves are strongly linked to possible others. The human self and, consequently, all the possible selves emerge from specific environments characterized by different social relations and cultural

resources. Within the boundaries that every being finds initially traced from birth and then, during development, might start tracing by themselves, an infinitive number of actions and interactions can offer specific forms of agency and ways of relating to possible others and forms others have to relate to a possible self (see also Glăveanu 2018).

Starting from these imaginative explorations, we could "construct an evaluative landscape of possible acts and outcomes" of selves and others (according to Seligman et al. 2013, p. 120). Below, we will try to apply an imaginative exploration of our false-belief tasks used to understand the theory of mind.

Consider the classical Sally-Anne task. What possible acts and outcomes can we imagine? The story recounts that Sally placed a marble into her basket and left the scene. Anne takes the marble out of Sally's basket and hides it in her own basket. Sally returned and the key question is: Where will she look for her marble? The answer is that Sally believes that the marble is in her own basket. However, if we image several landscapes, different key questions and outcomes can come to mind. For example, consider some possible alternatives of Sally's thought process: (a) the two baskets are identical, Sally does not remember which is hers (the marble is in any case in one of the two baskets); (b) Sally knows that Anne is nosy, and she could think that Anne had looked in her basket and had taken her marble to observe it better, so now the marble could be in Anne's hand (both baskets are empty); (c) Sally knows that Anne is spiteful, she could think that Anne had moved her marble as a joke to the second basket (the marble is in the second basket); and so on. Endless different scenarios may be created by the imagination. The more different rereadings of the same scenario we are able to make, the better chance we have of understanding exactly what has happened, in particular when we refer to reality. Moreover, we can also consider all the possibilities of why Anne and Sally act out certain behavior. Why does Anne put her marble in the basket? Does she want to hide it, to keep it safe, to take it to someone or give it to them in the basket, and so on. Why does Sally move the marble to the second basket? Is it a prank, or to be spiteful (Sally was angry with Anne), or is Sally bored and moves the marble just to fill in the time, and so on.

The combination of all the possibilities of decisions, intentions, and even emotions can generate an infinite number of different representations of reality. Experiencing self-other different perspectives offers individuals new views from which to understand and act on the world with flexibility, creativity, and imagination. This experience can significantly enact, increase, and define social interactions and communication, also increasing the well-being of oneself and others.

Summary

In the last few years, we have extended more and more our knowledge about how ToM works, its development, its neural underpinnings, and also about how it is affected by specific clinical cases, such as autism.[1] However, as Rakoczy (2017) highlighted, many different questions remain open for the future. For example, how do we develop from early implicit ToM to later explicit ToM? What are the neurocognitive foundations of ToM reasoning in infants and adults, and how do they change during development?

Furthermore, if we consider the ToM in relation to the possible, new exciting questions arise. What happens if we train people to a divergent way of thinking that considers all the other possible behaviors or mental reasonings? Will their theory of mind change? Will their affective responses, such as empathic responses, and rational judgments change regarding other behaviors and reasonings? Moreover, if people learn to analyze all possible alternatives before acting, they could learn to choose the best alternative for themselves and others. Increasing the consciousness of the variety of situations that the possible may allow people to start seeing the world not as it is, but as it could be, increasing their willingness to improve it.

[1] For the sake of brevity, the relation between ToM and various other pathologies has not been included in this entry; for example, the importance of the ToM's deficits in autism has not been covered; see Baron-Cohen (2000), Fletcher-Watson and Happè (2019), Mitchell (1997).

Cross-References

- ▶ Creativity
- ▶ Imagination
- ▶ Insight

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970.

Astington, J. W., Harris, P. L., & Olson, D. R. (1988). *Developing theories of mind*. Cambridge: Cambridge University Press.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110–118.

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 233–251). Oxford: Blackwell.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford: Oxford University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

Bretherton, I., McNew, S., & Beeghley-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition* (pp. 333–373). Hillsdale: Lawrence Erlbaum.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187–192.

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science, 16*, 378–384.

Doherty, M. (2009). *Theory of mind*. Philadelphia: Psychology Press.

Erikson, M. G. (2007). The meaning of the future: Toward a more specific definition of possible selves. *Review of General Psychology, 11*(4), 348–358.

Fletcher-Watson, S., & Happè, F. (2019). *Autism: A new introduction to psychological theory and current debate*. New York: Routledge.

Fodor, J. A. (1978). Propositional attitudes. *The Monist, 61*, 501–523.

Frye, D., & Moore, C. (1991). *Children's theories of mind. Mental states and social understanding*. Hillsdale: Lawrence Erlbaum.

Glăveanu, V. P. (2018). The possible as a field of inquiry. *Europe's Journal of Psychology, 14*(3), 519–530.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science, 17*(5), 647–659.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.

Lewis, C., & Mitchell, P. (Eds.) (1994). Children's early understanding of mind: Origins and development. Hillsdale, NJ: Erlbaum.

MacNamara, J., Baker, E., & Olson, C. (1976). Four-year-olds' understanding of pretend, forget, and know: Evidence for propositional operations. *Child Development, 47*, 62–70. https://doi.org/10.2307/1128283.

Mitchell, P. (1997). *Introduction to theory of mind: Children, autism and apes*. London: Edward Arnold Publishers.

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition, 39*, 107–127.

Moore, C. (1996). Theories of mind in infancy. *British Journal of Developmental Psychology, 14*, 19–40.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258.

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*, 59–74.

Rakoczy, H. (2017). Theory of mind. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge encyclopedia of child development* (pp. 505–512). Cambridge, UK: Cambridge University Press.

Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science, 8*(2), 119–141.

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? *Psychological Science, 22*(7), 878–880.

Shultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology, 17*, 456–471.

Shultz, T. R., Wells, D., & Sarda, M. (1980). The development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social & Clinical Psychology, 19*, 301–310.

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology, 44*(6), 1789–1795.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587–592.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*(1), 30–44.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580–586.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives, 11*, 207–214.
Wellman, H. M., Cross, D., & Watson, J. (2001). Metaanalysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.
Whiten, A. (1991). *Natural theories of mind*. Oxford: Basil Blackwell.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.
Zittoun, T., & de Saint-Laurent, C. (2015). Life-creativity: Imagining one's life. In V. P. Glăveanu, A. Gillespie, & J. Valsiner (Eds.), *Rethinking creativity: Contributions from cultural psychology* (pp. 58–75). London: Routledge.

Thought Experiments

Michael T. Stuart
Department of Philosophy, University of Geneva, Geneva, Switzerland

Keywords

Thought experiment · Imagination · Possibility · Epistemology of imagination

Overview

Thought experiments – like Schrödinger's cat and the trolley problem – are a way for inquirers to focus the power of the imagination. What makes a thought experiment different from fantasies and daydreams is that they aim to produce new knowledge, wisdom, understanding, illumination, or something like that. They typically also have a narrative structure, with a beginning, middle, and end. Usually there are several phases in a thought experiment: one in which we set up some imaginary scenario, another in which we "see" what happens in that scenario, and, finally, one in which we draw some conclusions. At this level of description, thought experiments are like laboratory experiments, except they are carried out in the imagination.

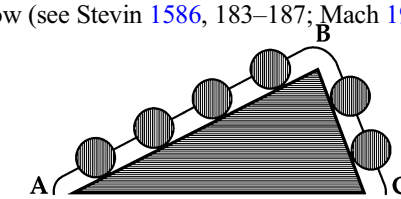
This entry will consider what thought experiments are, who performs them, how they have been investigated, what they aim to do, how they work, and how they connect to the possible.

What is it to be happy? Perhaps being happy is just feeling pleasure, like resting your legs after a long day's work, or listening to a favorite song. Robert Nozick presents a thought experiment to test this view (Nozick 1974). Suppose there was a machine you could enter, which would attach itself to your brain, and stimulate it so that you felt you were experiencing all the pleasures you've always dreamed of: the best food, fame, meaningful work, true love, etc. You would have no memories of entering the machine or of your previous life, and you must enter the machine for the rest of your life or not at all. Would you?

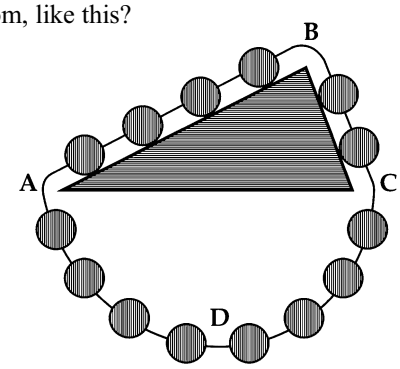
If pleasure is all there is to a happy life, we should all want to enter the machine. But the majority of people would refuse (Hindriks and Douven 2018). Why? Perhaps it is because there is more to happiness than pleasure. Maybe connections to real events and people matter too.

One thing that makes this a thought experiment is that when we begin, we don't know what will happen. We use our imagination, and we learn something new.

Here is another. Imagine a frictionless triangular prism with a chain draped over it, as in the picture below (see Stevin 1586, 183–187; Mach 1905).



How will the chain behave? Perhaps we think the chain will slide toward A or toward C. Okay. But now, what if the chain is connected around the bottom, like this?



Well, in this case, if the chain slides toward A, it will slide that way forever, and we will have a