Reasoning about self and others

Meijering, Ben

Chapter 1

Introduction

Theory of mind: Understanding others

We live in a social world in which we frequently interact with others. In our jobs, for example, we may collaborate with colleagues and negotiate with superiors, and in our leisure time we may compete with friends when playing a board game. These different scenarios have in common that in each of them we are trying to understand one another. For example, to collaborate with a colleague we need to know what her goals are to be able to find commonalities, as a basis for working together. To compete with a friend in a board game we need to know what her goals are to be able to anticipate her actions. In either case, we are reasoning about goals, which happen to be intangible, hidden to the eye. In spite of this characteristic of goals, human beings are quite proficient at inferring goals and other so-called mental states such as beliefs, desires, and intentions. In fact, the ability to infer mental states of others seems to be unique to human intelligence (Call & Tomasello, 2008). This ability has often been referred to as theory of mind (Onishi & Baillargeon, 2005; Premack & Woodruff, 1978; Wimmer & Perner, 1983), and because the ability to infer mental states of others is pivotal to human nature, many studies investigate the nature and development of theory of mind (Wellman, Cross, & Watson, 2001).

Studying theory of mind

The most influential paradigm to investigate theory of mind is the false-belief task, also known as the Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). In this task, children listen to a story with two characters: Sally and Anne. Sally is playing with a marble, and before she leaves the room she places the marble in her basket. In Sally's absence, Anne moves the marble and hides it in another location, a box. Then, Sally returns and children are asked where she will look for her marble. To pass this task, children need to understand that Sally will look for the marble in the basket, whereas in fact the marble is now in the box.

Some developmental studies have shown that infants as young as 15 months are already susceptible to the mental states of others. Onishi and Baillargeon (2005), for example, have shown that 15-month-old infants are surprised when their expectation that an agent will act according to her beliefs is violated. In Onishi and Baillargeon's false belief task, children saw that an agent, who believed that a slice of watermelon was stored in a box, was unaware of the watermelon being moved to a second box. When the agent reached for the watermelon, after it had been moved, the children looked longer at the scene if the agent reached for the second box instead of the first. The children looked longer because the agent did not act according to her false belief that the watermelon was hidden in the first box. Importantly, Onishi and Baillargeon administered an implicit test of theory of mind, using the violation of expectation paradigm, which did not require the infants to explicitly reason about the agent's mental states. As the violation of expectation paradigm does not require explicit reasoning, the infants' understanding of mental states would not be obscured by cognitive functions that are yet to be developed.

By the age of 4, children have sufficiently developed cognitive functions to pass the original false belief task (Wellman et al., 2001; Wimmer & Perner, 1983). This task requires a deliberate

response, as children are explicitly asked where the agent will look for the toy. A deliberate response is an additional step that involves effortful processing, which is clearly illustrated by a dissociation between two distinct measures of theory of mind within the same sample (Clements & Perner, 1994; Ruffman, Garnham, Import, & Connolly, 2001). In Ruffman et al.'s study, for example, children looked at the correct location, where the agent falsely believed the toy should be, but answered incorrectly that the agent would look at the toy's actual location. To this day, this remarkable finding still does not have one succinct explanation, and there are many studies investigating which distinct processes constitute the developmental trajectory of theory of mind.

Roughly speaking, there are two clearly opposing camps that have different ideas about which processes are involved in the development of theory of mind. One camp argues that children's understanding of mental states undergoes a conceptual change (Gopnik & Wellman, 1992). Not until children understand that, for example, beliefs do not need to reflect reality, do they understand that others can have beliefs that differ from their own. The other camp argues that children first need to develop other important cognitive functions that, once sufficiently developed, will enable inference of mental states (e.g., Carlson, Moses, & Breton, 2002; Sabbagh, Xu, Carlson, Moses, & Kang, 2006). These cognitive functions are required to support the costly computational processes that are involved in mental state inference. Of course, children still need to be able to discern mental states, for example, by means of a so-called theory of mind mechanism (e.g., Leslie, Friedman, & German, 2004), or a minimal understanding of concepts in general. Accordingly, there seems to be merit in hybrid theories as well, combining the theories of the two camps delineated above, as application of theory of mind seems to be an effortful process that requires sufficiently complex mental state representations.

Given that application of theory of mind is a costly computational process, it may not be surprising that it takes another two years before children, of 5 to 6 years old, start understanding that others, too, apply theory of mind (e.g., Miller, 2009). They learn to understand recursive structures such as "John thinks that Mary thinks that…" (Perner & Wimmer, 1985). Comprehension of such structures requires application of second-order theory of mind, which is a challenge for both children and adults (Flobbe, Verbrugge, Hendriks, & Krämer, 2008; Hedden & Zhang, 2002; Perner & Wimmer, 1985; Qureshi, Apperly, & Samson, 2010; Raijmakers, Mandell, Van Es, & Counihan, 2013). Application of higher orders of theory of mind would exceed the cognitive resources of most people. The dynamics of declarative and procedural memory, for example, would probably not support such complex reasoning. Fortunately, in many circumstances second-order theory of mind seems to be the highest level of theory of mind that is still advantageous for people to use (De Weerd, Verbrugge, & Verheij, 2013).

Contribution of this dissertation

This thesis describes a detailed investigation of theory of mind in adults, who have sufficient declarative and procedural knowledge of mental states to interpret the behavior of self and others, in contrasts to infants and children. The thesis therefore contributes to the field in at least one obvious way: As adult understanding of mental states has undergone the earlier

proposed conceptual changes (Gopnik & Wellman, 1992), the findings reported here can be interpreted as a measure of the cognitive functions and resources required to apply theory of mind. In other words, the findings in this thesis help shed light on the procedural building blocks that constitute inference of mental states. A better understanding of the procedural building blocks is important, because it may be of help in both the clinical and practical domains.

Another contribution of this thesis is that it shows how two-player games may be a good alternative to test for theory of mind. In contrast to false-belief stories, games can be presented many times and in many different configurations to the same individuals. The assumption that these games are interpreted in terms of mental states remains implicit in some chapters (Chapters 2, 5, and 6). However, Chapters 3 and 4 explicitly test whether two-player games require a theory of mind. It turns out that these games are especially successful in varying demands on mental state reasoning.

The thesis starts with a broad question in Chapter 2. The question is whether theory of mind is a fixed skill, or an ability that is susceptible to improvement by means of supportive measures that help structure inference of mental states. Chapters 3 and 4 detail an investigation into the computational costs associated with taking either one's own or someone else's perspective. Most definitions of theory of mind state that it is an understanding of mental states of self and others. However, few studies seem to investigate the difference between taking one's own and someone else's perspective. Chapters 5 and 6 provide a detailed analysis of particular strategies that people use when they are reasoning about others' mental states. These studies show that people rather use simple strategies that may not be optimal but yield correct responses in most cases. Only when they really have to, do people invest their resources into forming complex mental state representations. In other words, people may have the ability to infer complex mental states, but still fail or choose not to use it.