Learning to reason about other people's minds

Lisette Mol, Rineke Verbrugge, Petra Hendriks
Institute of Artificial Intelligence, University of Groningen

31st October 2004

Abstract

To investigate to what extent people use and acquire complex skills and strategies in the domains of reasoning about others and natural language use, we conducted an experiment in which it was beneficial for participants to have a mental model of their opponent, and to be aware of pragmatic inferences. We found that, although participants did not seem to acquire complex skills during the experiment, some participants made use of advanced cognitive skills.

1 Introduction

In every day life, people frequently make use of their ability to reason about others and to infer the implicit meaning of sentences. Consider the following two situations:

Situation 1 You are called by a friend who asks you for a phone number. You know the number by heart, so you ask her whether she has pen and paper. She answers you with "No, I don't". Can you conclude that she also does not have a pencil and paper ready?

Situation 2 You are playing happy families and you are the first to pose a question. You ask your opponent for the mother of a certain family. Your opponent replies with "not at home". Can you conclude that he doesn't have any members of this family?

In the first case, you know that your friend has the desire to be cooperative and thus your reasoning would be something like, 'She does not have a pencil, for if she did she would have told me so, since she knows it is relevant'. In the second case you know that your opponent does not want you to know which cards he has, since he has the desire to win the game. You therefore are aware that he would not tell you whether he has any other members of the family, unless he really had to, and thus you do not conclude that he does not have them.

These examples make it clear that, to successfully interact with people, agents will need advanced cognitive skills like reasoning about others and drawing pragmatic inferences, and will need to know when to use these skills. It would therefore be interesting to know how humans acquire and use such skills. In the study described below, we have investigated to what extent people use and acquire complex skills in the domains of reasoning about others and language use.

2 Background

Theory of mind use

One of the advanced skills that we are interested in is Theory of Mind (ToM) use. Although children from the age of six are able to distinguish between their own mental states and those of others, Boaz Keysar, Shuong Lin and Dale J. Barr [3] argue that even adults do not reliably use this sophisticated ability to interpret the actions of others. They found a stark dissociation between the ability to reflectively distinguish one's own beliefs from others', and the routine deployment of this ability in interpreting the actions of others. The second didn't take place in their experiment. In other experiments by the same research group, similar results were found [4], [5], [6].

To have a first order ToM is to assume that someone's beliefs, thoughts and desires influence one's behavior. A first-order thought could be: 'He does not know that his book is on the table'. In a second-order ToM it is also recognized that to predict others' behavior, the desires and beliefs that they have of one's self and the predictions of oneself by others must be taken into account. So, for example, you can realize that what someone expects you to do will affect his behavior. A

second-order thought could be: 'He does not know that I know his book is on the table'. To have a third order ToM is to assume others to have a second order ToM, etc.

Pragmatic inferences

Another, but related skill that we have investigated is language use, especially drawing pragmatic inferences. According to Grice [2], people use the quantity maxim to infer the implicit meaning of a sentence. The quantity maxim states that interlocutors should be as informative as is required, yet not more informative than necessary. Using this maxim it can be inferred that if someone says 'Some Italians like ice-cream', it is the case that not all Italians like ice-cream. This is because if all Italians would like ice-cream, the speaker would probably have known this, and thus would have used the more informative term *all* instead of the weaker term *some*, since otherwise the quantity maxim would have been violated.

Some and *all* are scalar terms. Scalar terms can be ordered on a scale of pragmatic strength. An example is $\langle a, some, most, all \rangle$ which is ordered from weak to strong.

In [1], Feeney et al. propose that there are three stages to people's understanding of *some*. The first is the logical interpretation which precedes children's sensitivity to scalar implicatures. The second stage is the pragmatic interpretation which results from drawing pragmatic inferences. This is in line with the results found by Noveck [8] and Papafragou & Musolino [9]. Feeney et al. found evidence for a third stage, in which adults can choose a logical interpretation over a pragmatic interpretation, even though they can make the pragmatic inference that *some* implies *not all*.

3 Research Question and Hypotheses

The context described in the previous section led us to the following research question: *'To what extent can people acquire and use complex skills and strategies, in the domains of reasoning about others and language use.'* We narrowed this down to the specific case of playing Master(s)Mind(s) [7]. This is a symmetric version of the usual Mastermind, of which a variant was used in our experiment, which is described below. To find an answer to the research question, the following three hypotheses were stated:

Hypothesis 1 *Performing a task and simultaneously reflecting upon this task can be seen as a form of dual tasking.*

Hypothesis 2 *In an uncooperative conversation, people will shift their interpretation and production of quantifiers from a pragmatic (using Grice's quantity maxim) to a less pragmatic (not using Grice's quantity maxim) use.*

The idea behind this hypothesis is that in an uncooperative situation, people will be aware that others are trying to reveal little information (first order knowledge) and therefore will be aware that the quantity maxim does not hold. In addition, people may develop more logical productions to be less informative themselves.

Hypothesis 3 *In using quantifiers, people make use of an automatic process, which results in a pragmatic use of the quantifier. This automatic process can be 'overruled' by a deliberate reasoning process, which results in a logical use of the quantifier.*

4 Experiment

Participants (native Dutch speakers) had to complete two sessions, each about three hours, in which they played a symmetric head to head game. In this game they had to correctly guess the secret code, consisting of four different, ordered colors, of their opponent. Players gave each other feedback by selecting Dutch sentences, which differed in pragmatic strength, from a list. The game was about gaining as much information as possible, while at the same time revealing as little information as possible. Because of this second aspect, the conversation is not fully cooperative and thus hypothesis 2 is relevant.

During the game, players had to submit their interpretation of the sentences they received as feedback, through a code. For each right color in the right position they had to select a black circle and for each color which was correct but in the wrong place, a white circle. To represent ambiguity and vagueness, participants could submit more than one combination of black and white. Because the number of correct colors and correct positions was known to the experimenters, this gave insight in the production as well as the understanding of the sentences.

Let's look at an example. Imagine John having the secret code red, blue, green, yellow and Mary guessing red, orange, yellow, brown. The evaluation of this situation is that exactly one guessed color is right and in the right place and exactly one guessed color is right, but in the wrong place. John has to choose two feedback sentences to communicate to Mary, one about color and one about position. He could say 'Some colors are right.' and 'There is a color which is in the right place.' This would indicate that John thinks that *some* can mean *exactly two* and that *a* can mean *exactly one*. This is a pragmatic production (in accordance with Grice's maxims). If he had chosen the sentence 'One color is right.', then he would allow *One* to mean *exactly two*. This would be a more logical production (in logic *One* is true in case of *at least one*).

Mary now has to give her interpretation of the sentences chosen by John. So if she thinks that, given the first two sentences it could be the case that two colors are right, of which one is in the right position, she would submit black, white as a possible interpretation. If she considers the situation where three colors are right, of which two colors are in the right position possible as well, she would also submit black, black, white. If she would only submit the first possibility, her understanding would be pragmatic. If she would also submit the second case, her interpretation would be more logical.

In the experiment Mary would have to give John feedback about her guess compared to her secret code as well, and John would then submit his interpretation of those sentences. Each turn, one player can make a guess, in this example Mary.

During the experiment participants had to answer questions. The purpose of those questions was to get information on their strategy and the order of the theory of mind they were using. For the same purpose, participants completed a questionary after each session.

5 Results

Three out of twelve participants showed clear signs of the use of second order ToM. One additional participant probably used second order ToM as well, but in this case it was less clear. An example of second order ToM use in this game is that agent 1 assumes that the guesses made by agent 2 are evasive about agent 2's own code, since agent 2 does not want agent 1 to know agent 2's secret code. All of these four participants played in accordance with a strategy of being uninformative and had a fairly to strict logical language use.

The remaining eight participants all used first order ToM. An example of first order ToM use in this game is to take into account what the opponent already knows. Two of them had a strategy of being uninformative and a fairly logical language use. The other six used the strategy of being informative or a strategy which did not consider the amount of information being revealed and had a fairly to strict pragmatic language use.

All participants with a strategy of being uninformative and a fairly to strict logical language use showed a type of behavior which the others did not show. This behavior consists of, for example, favoring sentence 1 over sentence 2 in a case where, from a logical perspective, they both hold.

1. 'Some colors are right.'
2. 'All colors are right.'

All participants which used second order ToM did so from the start. One participant shifted from a fairly pragmatic to a fairly logical use, and had a strategy of being uninformative. Three participants shifted from a fairly pragmatic to a fully pragmatic use, they did not use a strategy of being uninformative. The other participants were constant in their language use. One participant shifted from a strategy of being informative to a strategy of being uninformative, he had a fairly logical language use. One participant even abandoned the strategy of being uninformative, to give the opponent a better chance of winning, and had a fairly pragmatic use of language.

The participants using more advanced strategies, clearly had to put little effort into playing the game and understanding the computer program used. The people with the least advanced strategies made more mistakes in playing the game than others.

6 Discussion and Conclusion

Although no evidence was found that people learned to do so during the experiment, some participants clearly showed the use of complex skills as meant in the research question.

They were able to use an advanced strategy by applying their ToM. It could very well be the case that their understanding of the uncooperative situation, attributed to their logical use of the quantifiers.

Although the evidence found for hypothesis 1 is not very convincing, there is no reason to abandon hypothesis 1 because of this experiment. Hypothesis 2 should be abandoned. It is clear that it does not hold for most participants: The shift does not occur during the experiment but either at the start, or not at all. Even though this hypothesis does not hold, the uncooperative situation may very well influence language use, at least for some people.

Since some people shifted from a fairly pragmatic to a fully pragmatic use of quantifiers, it seems like a pragmatic use does not result from a fully automated process. Hence, hypothesis 3 should be abandoned. It could very well be the case that the people who's language use was more logical are also capable of a more pragmatic language use in daily life. Since some participants used a fairly to fully logical language use consistently from the start, it does not seem like they have to put a lot of effort into overruling a process.

7 Future Work

In the light of earlier work, for example [3], it would be interesting to investigate whether people make mistakes in the use of first and second order ToM while playing Master(s)-Mind(s). It was already shown that people make mistakes in first order ToM in other experiments.

It would also be very interesting to investigate what kind of language use, pragmatic or logical, the subjects with the more advanced strategies use in cooperative situations. This could make it more clear whether the process of drawing pragmatic inferences is overruled or not, and what the influence of uncooperative situations on language use is.

Knowledge of ToM and language use would be very useful in designing conversational agents, because if humans draw inferences differently, depending on the nature of the situation, artificial agents should also do so, and should be able to take into account that others may do so.

References

[1] A. Feeney, S. Scrafton, A. Duckworth, S.J. Handley. The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology.*

[2] P. Grice. *Studies in the Way of Words.* Cambridge, MA: Harvard University Press, 1989.

[3] B. Keysar, S. Lin, D.J. Barr. Limits on theory of mind use in adults. *Cognition* (89): 25-41, Elsevier Science B.V., 2003.

[4] B. Keysar, D.J. Barr, J.A. Balin, J.S. Brauner. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* Volume 11 Number 1, American Psychological Society, 2000.

[5] B. Keysar, D. J. Barr, A. Balin, T.S. Peak. Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language* (39): 1-20, Academic Presss, 1998.

[6] W. S. Horton, B. Keysar. When do speakers take into account common ground? *Cognition* (59): 91-117, Elsevier Science B.V., 1996.

[7] B. Kooi. Master(s)mind(s). Unpublished manuscript. See http://www.philos.rug.nl/~barteld/.

[8] I. Noveck. When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* (78): 165-188, Elsevier Science B.V., 2001.

[9] A. Papafragou, J. Musolino. Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*(86): 253-282, Elsevier Science B.V., 2003.