Canteen Dilemma Article

Thomas S. Nicolet

April 30, 2019

Abstract

Social interaction and cooperation often requires us to reason about the beliefs of others in order to predict their behavior. The canteen dilemma is a coordination game with imperfect information intended to test how people behave and reason in circumstances where they lack crucial knowledge. Our findings indicate signs of first and second-order theory of mind being used but not more. Most importantly, the higher orders of theory of mind are indicated by participants estimates of success, which comes in varying degrees. In other words, we argue that our results show that higher order-theory of mind can be exhibited in continuous degrees and not just be described by having a n or $n+1$ theory of mind.

Contents

1 Introduction

Understanding others in terms of their beliefs, intentions or desires is essential for intelligent social interaction. The cognitive capacity to understand and reason about such otherwise unobservable mental states of others is often referred to as theory of mind, social cognition or perspective taking. Higher order theory of mind is the ability to attribute theory of mind reasoning to others. This article is concerned with the limits of this capacity, how robust it is and whether the capacity is discontinuous in terms of being able to reason up to n order of theory of mind but not $n+1$. We conducted an experiment where participants played a game called the canteen dilemma. The game involved two players, each assigned an arrival time and told they arrived 10 minutes apart. It requires consideration of not just the other player's possible arrival times, but also what

the other player considered to be your possible arrival times. Through the certainty estimates included in our experiment, our findings show that the mental capacity of having a higher order theory of mind might comes in continuous degrees rather than discrete orders.

The capacity of having a theory of mind develops through early childhood and is essential for recognizing that others have minds of their own with distinct desires, intentions and beliefs. As such, this capacity allows humans to both interpret the behavior of others and infer underlying mental traits, while also reasoning and predicting their future behavior based on their assumed mental states. Acknowledging that other organisms are sentient like ourselves enables us to distinguish between them and physical constructs. While this is essential for moral considerations as well, this aspect is not the focus of the present article, but it does indicate the important social function of having a theory of mind.

Much of the literature on theory of mind has been on the developmental aspect and limitations in relation to having a higher order theory of mind. Studies on child development indicate that children acquire a first-order theory of mind between ages 3 to 5 [76] and a second-order at ages 6 to 8 [60]. Some argue that this capacity is present even earlier [15, 49] and some argue that basic theory of mind aspects as goal perception are present already in 9 to 12-month-old infants [20]. This is relevant because it indicates some of the ease of making mental constructs. In other words, having a theory of mind does not necessitate a deliberate reasoning process. For example, seeing a person raise their hand in a classroom is unlikely to be seen first as mechanical movement which is then interpreted, since people are so conditioned to recognize others as intentional beings.

Adults are usually limited at most second-order theory of mind. There is evidence that theory of mind is not just limited in the sense of only having a second-order theory of mind but also that any application of theory of mind might be severely when it comes to spontaneous use. See for example Lin et al. [50] for evidence that when interpreting the actions of others, the default is to rely on one's own mental states and beliefs as representative of the mental states of others. Lin et al. argue that effortful attention is required in order to use what is known about other's beliefs. Birch & Bloom [11] also show a *curse of knowledge* bias in the sense that adults' own knowledge about an event can make them worse at correctly attributing false beliefs about that event to others. Keysar et al. [47] show a disassociation between adult's ability to reflectively distinguish between their own beliefs and others and actively utilizing this ability when interpreting the actions of others.

1.1 Problem of positing discontinuous bounds on social cognition

There is a possible pitfall when discussing the limits on theory of mind reasoning. It is tempting to assume that there is a fixed discontinuous bound on social cognition, or in other words that people can reason up to n order of theory of mind but not $n+1$. If a person is able to apply first-order theory of mind and model the mental state of another person, it seems natural that this leads to reasoning about that person doing the same to you, effectively applying a second-order theory of mind. But this line of reasoning can not go on for ever, and as we have seen, it often stops after two iterations. But it is conceivable that there is greater variation than simply n or $n+1$ orders of theory of mind and such variation is important for our understanding of theory of mind. In other words, describing orders of theory of mind discretely might be an approximation which does not give a proper picture of the real theory of mind which is used in everyday circumstances. Our findings show variation within those exhibiting n order theory of mind such that some are closer to $n+1$ than others. It indicates that (higher) orders of theory of mind should be understood as a continuous capacity, that is, as there being variying capacities between some n and $n+1$ order of theory of mind.

2 The canteen dilemma experiment

The canteen dilemma is a two-player coordination game with imperfect information. The game has a structural affinity to the consecutive number example in Ditmarsch & Kooi [22]. It is framed in a thematic story as to make some of the logical reasoning easier [55, 73]. The story is the following. Each player is told that they and their colleague arrive for work every morning between 8:00 am and 9:10 am. They always arrive 10 minutes apart but only know their own arrival time. The payoff structure is ordered such that $(1 > 2 > 3)$ where (1) going to the canteen together if both arrive before 9:00 am, (2) going to the office together at any time and (3) all other configurations, that is, discoordination or either player going to the canteen at 9:00 am or later. The game consists of a numbers of rounds where each player are given their own arrival time and have to decide between going to the canteen or the office. The structure of the game entails the unintuitive result that if a person decides to go to the office at some time t because the other player might go to the office if they arrive 10 minutes later, and if the first person assumes the other player is like herself, then they will never go to the canteen.

We hypothesize that being able to reason about and predict the actions of the other player is facilitated by the participants theory of mind. As higher-order theory of mind is limited, and in any regard not common knowledge among participants, we do not expect players to play the optimal office-only strategy. Furthermore, since recognizing that one ought to go to the office in order to avoid discoordination depends on a capacity for more than 2 iterations of higher-order theory of mind, we expect that even participants who exhibit some order of theory of mind still believe that their strategy is safe as long as they move their *cutoff* for going to the canteen to a sufficiently early time.

2.1 Method and design

2.1.1 Participants

Our experiment included 714 adults on Amazon's Mechanical Turk (AMT) platform. Certain settings were applied in order to only include participants from Canada or the United States, participants with at least 500 approved Human Intelligence Tasks (HIT's) and a HIT approval rating of at least 98%. Participants were also given a unique ID such that they could only enter the experiment once and were awarded a $2 participation fee if they completed the HIT. We also conducted two experiments at the Technical University of Denmark with 106 and 50 participants each during coursework.

2.1.2 Materials

The main experiment was conducted on Amazon Mechanical Turk (AMT) which is an online crowdsourcing platform. The experimental setup was implemented in oTree 2.1.35 software [17]. AMT works as an online labor market where workers (also called turkers) can perform HIT's for monetary compensation. The platform has been used by social and economic researchers in lieu of typical lab experiments with local university students. Experiments on AMT have been shown to live up to the standards set by other data collection methods [10][13] and to provide reliable, replicable and more diverse data than legacy methods using university students [19][22][43][53][63].

After accepting our HIT and providing informed consent, participants were put in a 'waiting room' until they were paired up with another participant. After a group was formed, participants were directed to an initial introduction page which detailed the rules of the game (see Appendix A for screenshot). After reading the instructions, participants were directed to round 1 (of 10) where they were given their own arrival time

and asked to make a decision between between going to the canteen or the office. After making this decision they were prompted to estimate how certain they were that the other player made the same choice as them, ranging from very uncertain, slightly certain, somewhat certain, quite certain to very certain. After both players made their choices, they were prompted to a results page showing them the results of the previous rounds, including arrival times for both players, their choices, their own certainty estimate and resulting payoff.

Players payoff were implemented using logarithmic scoring as a proper scoring rule. Payoffs were distributed by initially awarding each player $10 which was reduced each round depending on well they did and their certainty estimate of success. Research literature on eliciting belief also show that forecasts elicited from observers through proper scoring rules are significantly more accurate and calibrated then those elicited from observers using an improper scoring rule. Calibrated is defined as: "a set of probabilistic predictions are *calibrated* if p percent of all predictions reported at probability p are true" [65]. There is also evidence that forecasts elicited by the logarithmic scoring rule seem to have significantly less dispersion than quadratic scoring rules even though both are proper scoring rules [58], which among other reasons favored the logarithmic scoring rule over the quadratic.

The experiments included four post-game questions and the DTU trials included 3 further questions (See appendix B)

2.2 Results

The first graph below shows percentage of participants choosing canteen at specific arrival times (blue line) and how many percent of those choosing canteen answered 'very certain' that the other player also chose canteen (orange line). Shaded areas represent 0.95 confidence intervals.

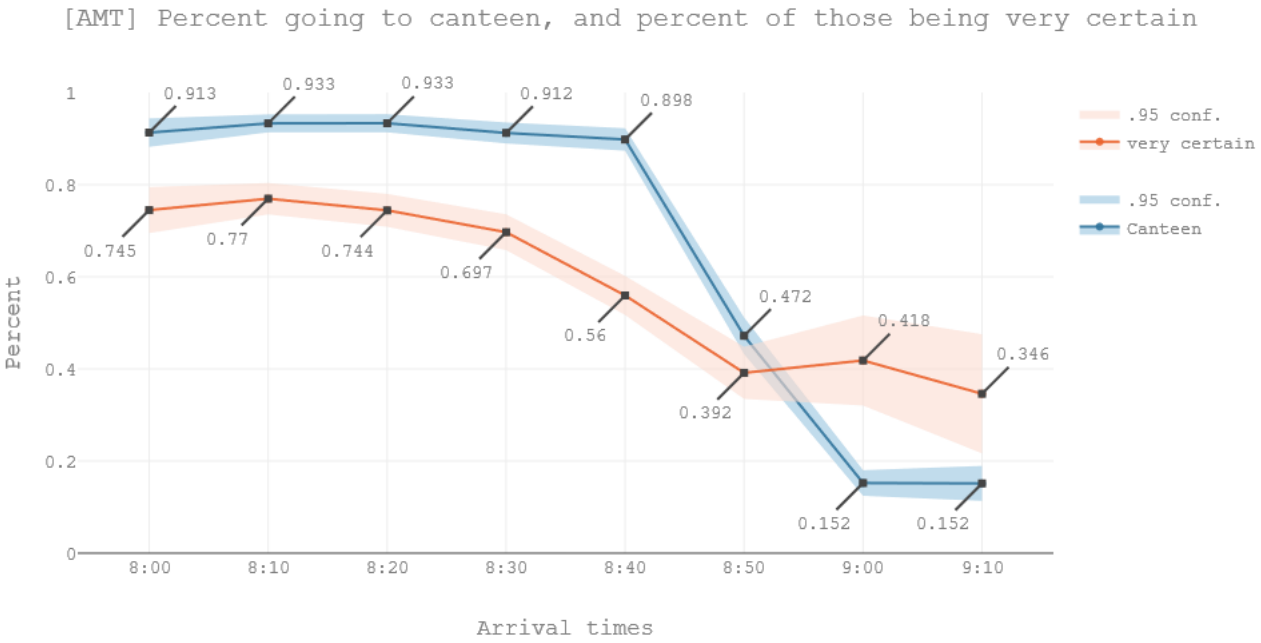


Figure 1. Canteen/office choices and certainty estimates.

Figure 1 shows that participants generally went to the canteen from 8:00 to 8:40 and to the office at 9:00 at 9:10, while office was slightly favored at 8:50. While we see a rather steady trend of canteen choices from 8:00 to 8:40, the percent being 'very certain' that the other player does the same drops from 8:10 to 8:40. The reason for going to the office at 8:50 is stated in the rules as "if you or your colleague arrives at 9:00 am or after, you should go straight to your offices" combined with the statement that the two players always arrive 10 minutes apart. In other words, going to the office at 8:50 arguably does not require participants to use their theory of mind. So we can postulate that going to the office at 8:50 requires 0-order theory of mind, 8:40 requires first-order theory of mind, 8:30 second-order theory of mind and so on.

While the general canteen choices at 8:00 to 8:40 seems to indicate a lack of any theory of mind, their certainty estimates tells another story. The percent being 'very certain' that the other person also chose canteen drops significantly from 8:00 to 8:40 (74.5% to 56%) while the percent going to the canteen only drops from 91.3% to 89.8%. This result indicates that participants' certainty about choices were indicative of a higher-order theory of mind. It is possible that participants were employing their higher order theory of mind but were unsure if the other player were doing the same, leading to canteen choices regardless.

2.2.1 Question regarding cutoff

Our results also seem to show that participants believed that as long as they arrived early enough, they could be sure to coordinate in the canteen. Participants were asked after the game ended: "Assume you could decide a cut-off point with your colleague prior to the game, meaning that you would go to the office at this time or later, and the canteen if arriving earlier. If you always had to be very certain about your decision, what time would you select?".

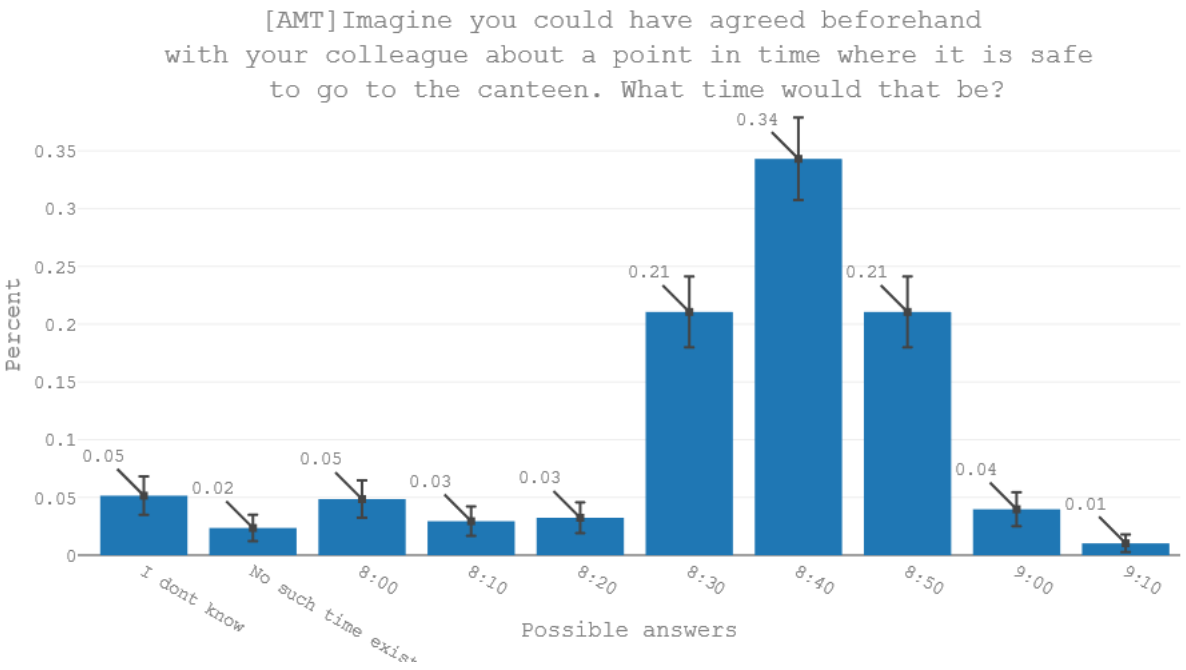


Figure 2. Bar-chart for a post-game question, error bars indicate 0.95 confidence intervals

Figure 2 shows a bar-chart for answers to the question about when it is the safe to go to the canteen. Most participants answers range from 8:30 to 8:50 (76% of all answers). The rest are somewhat evenly scattered among other answer possibilities, possibly due to random choices. Due to the pragmatics of language, we assume that an answer like 8:30 entails the belief that all earlier arrival times would also be deemed safe. The answer 8:30 (given by 21%) is likely due to the recognition that some might go to the office at 8:50, thereby making it unsafe to go to the canteen at 8:40. But if it is unsafe to go to the canteen at 8:40, some are bound to go to the office, which would mean 8:30 would not be safe either. This is not factored into the 8:30 answer. There are two possible reasons for this. Either those answering 8:30 do not believe that other players reason like they do or else they simply do not continue their theory of mind line of reasoning any further.

2.2.2 Question concerning curse of knowledge

The supplementary experiments at DTU (see Appendix C for plot corresponding to Figure 1) included the question "Imagine you arrived at 9:00//8:40 and you have been secretly informed that your colleague's arrival time is 8:50. Where do you think your colleague will go?" Half of the participants were given 8:40 as their own arrival time while the other half were given 9:00. They had to answer in both cases what they thought the other player would do. The question concerns whether player's own knowledge of their arrival time affects their prediction of the other player's decision. It relates to the curse of knowledge from Birch & Bloom [11] since participants might attribute their own belief (that it is early enough or too late to go to the canteen) to the other player. This is indeed the result we got in the first DTU trial.

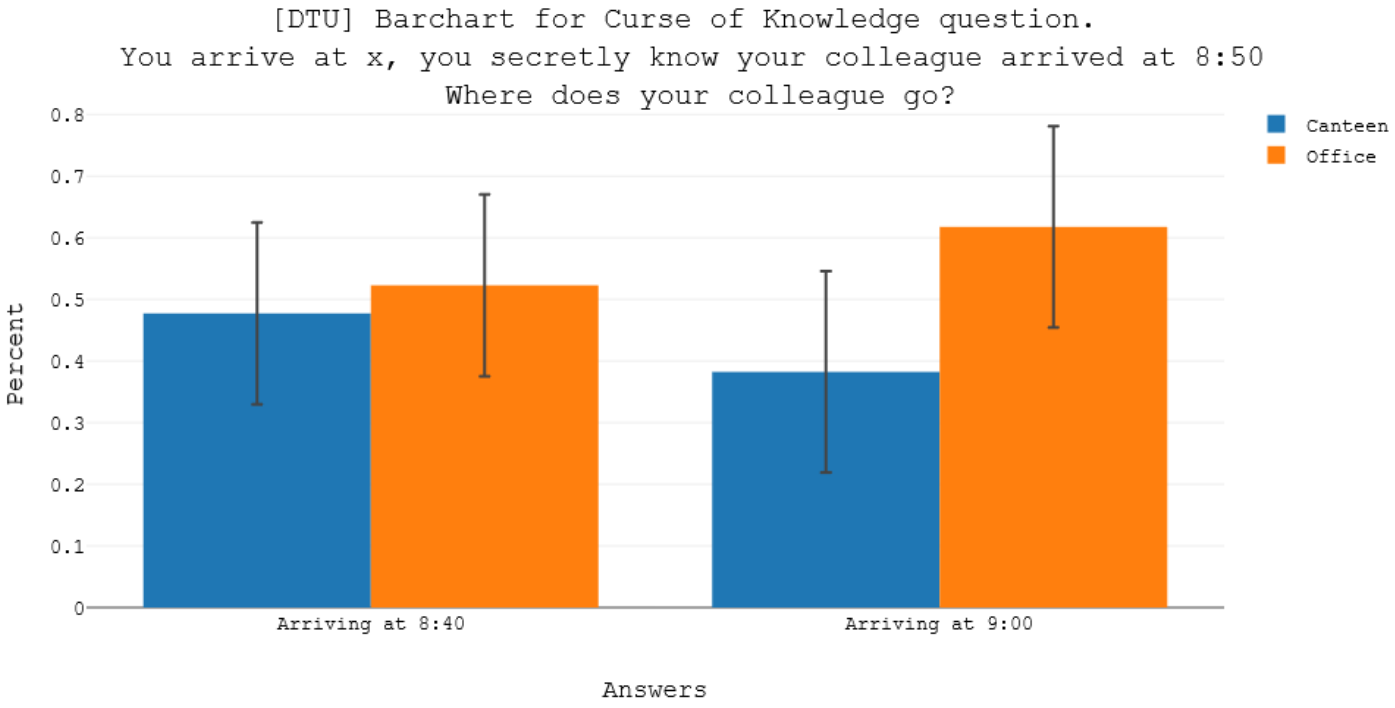


Figure 3. Grouped bar-chart for curse of knowledge question with 0.95 confidence intervals.

As figure 3 shows, participants who arrive at 9:00 are significantly more inclined to answer that the other player will go to the office than those arriving at 8:40. The confidence intervals are overlapping however and later results at at second DTU trial had differing results (see Appendix B). The data from the first DTU trial overall indicated a better understanding of the game than both of the other trials. That is, the second DTU trial had results which implied that participants had both understood the rules of the games better and employed more theory of mind in the other trials. None the less, the curse of knowledge seemed most prevalent in that case.

2.2.3 Categorizing free text answers [proper categorization in progress]

Participants were asked about what strategy they had used and there seemed to be some patterns in these answers. Some answer 'common sense' or 'intuition', while some simply stated that the earlier the arrival time, the more they went to the canteen, without further specification. A few answers were explicitly referring to first or second-order theory of mind considerations. Some answers were also indicative of forward inductive reasoning or reactive strategies, that is, decisions depending on the previous actions of the other player. Many of these answers indicate a belief that if some strategy, like the reactive strategy, would lead to certain coordination if only deployed correctly (once you know what the other player does at specific times, you can coordinate) or that if only the other player also employs first or second-order social reasoning, coordination can be successful. However, regardless of knowing the other players strategy, if that strategy contains a

canteen choice, coordination cannot be ensured. See for example the following strategy answer from the first DTU trial:

> "It was easy for 8:00, 8:10 and 8:20. There I would go to the canteen based on the fact, that my friend would arrive at latest at 8:30, and then they would in worst case scenario think that I arrived at 8:40. Thus we would both go for coffee. In the case of 8:30, I would also go to the canteen, but I would not be very certain cause my friend might be there at 8:40 and think that I would be there at 8:50. They might think that IF I am there at 8:50, I would think that my friend is there at 9:00, and thus I might go to the office. Thus for 8:30 and above it is not certain. It would depend on the history of our decisions... "

While it explicitly refers to the participants using their second-order theory of mind, it indicates that the third-order theory of mind is not only not considered, it is deemed irrelevant. This is important because it does not just show, like other studies, that people have different cognitive capacities which can lead to suboptimal results. It indicates that people do not have sufficient introspection to predict such suboptimal results. The lack of third-order theory of mind will also be discussed below.

2.3 Discussion

There are a few intriguing aspects of Figure 1. Notice that there are 15% going to the canteen at 9:00 and 9:10 and 47% at 8:50, even though the rules state that if one or the other player arrives at 9:00 , they should go straight to the office. There are also around 10% going to the office at times 8:00 to 8:40. There are bound to be some random answers in such tests therefore intentional or unintentional wrong answers can account for some of this. We might also wonder if the rules were ambiguously stated such that participants believed that they could go to the canteen whenever they felt like it.

Given a few of the free text strategy answers, it is possible that some participants were biased towards the canteens for reasons not stated in the rules. In other words, their real world preference for coffee in the canteen over work in the office might have persuaded them to choose the canteen as much as possible. This means that the positive facilitative effect of providing a thematic narrative to a logical reasoning test [55, 73] might not conclusive. Everyday narratives can help at explaining abstract structures, but it might do so by prompting heuristics which might sometimes be helpful and sometimes detrimental to proper reasoning. Hoever when conducting the first trial at DTU, responses were significantly more robust than at AMT:

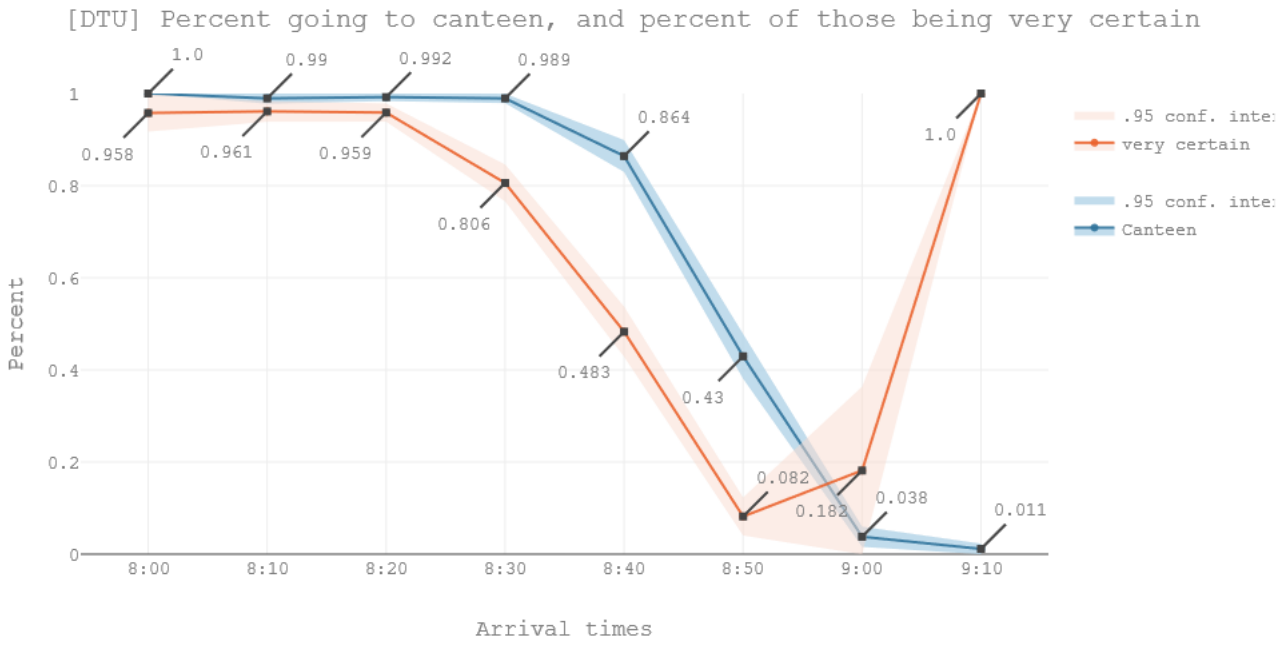


Figure 4. First DTU experiment.

Students in the first DTU trial made less office choices at 8:30 and earlier and less canteen choices at 9:00 and later. They also went relatively more to the office at 8:40 compared to earlier times. A plausible explanation for the difference in results is that the DTU students were more motivated to participate and understood the rules better and as such made less random choices. Therefore, a significant amount of the noise in the AMT results are likely due to random decisions. In the DTU trial depicted in Figure 4, we might still consider 43% a high amount of participants choosing the canteen at 8:50. At the second DTU trial, the wording of the instructions were changed from "if you arrive before 9:00 am, you have time to go to the canteen" to "if you *both* arrive before 9:00 am ...", emphasizing that both players have to arrive before 9:00 am to warrant a canteen choice. See Appendix D for appropriate plot. Contrary to expectations, this lead to a significant increase of canteen choices at 8:50, as 63% chose canteen at 8:50 in the second DTU trial. This implies that it was not the wording which lead to the behavior in question.

We assume that office choices at 8:40 and 8:30 are indicative of first and second-order theory of mind. Figure 4 shows that while DTU students might have chosen canteen in most of these cases, they are not entirely certain that the other player did so as well. We ascribe this uncertainty to first or second-order theory of mind respectively. However, like most established literature shows, there seems to be no signs of third-order theory of mind, since nearly all the DTU students went to the canteen at 8:20 while being 'very certain' that the other player did the same. Other results also indicate that going to the canteen at 8:20 was 'trivially' safe. This means that higher-orders of social cognition are not just excluded, they are deemed irrelevant. This makes sense for two reasons. First, recognizing the relevance of such higher orders of theory of mind requires considering the same orders of theory of mind. Secondly, since nearly every participant went to the canteen at 8:30 (without everyone being 'very certain' due to their second-order theory of mind), going to the canteen at 8:20 was indeed a safe choice.

More importantly however, is the possibility that the certainty estimates show the discontinuous nature of higher-order social reasoning. The certainty estimates were not binary but came in varying degrees, so if this certainty estimate can be indicative of a higher order theory of mind, it is also possible that higher order social cognition comes in continuous degrees as well. This deflects what Verbrugge calls the simplistic danger of positing fixed bounds on social cognition, such as 'everyone can reason up to n order of theory of mind, but not $n+1$'. It is likely that few actually believe that higher-order theory of mind actually comes in discrete degrees such that there is a sudden jump from first to second-order social reasoning and no middle ground. None the less, this is the picture often given in research on social cognition. This result arguably calls for a nuanced picture of the mental capacity related to representing the mental states of others which does not presume that it is an all or nothing capacity.

3 Conclusion

In conclusion, adults might have continuous degrees of higher orders of theory of mind and researchers should be careful when making approximations such as n or $n+1$ orders of theory of mind. This means that important cognitive resources might be missed in research which lumps together all those who exhibit a first-order theory of mind but not a second order theory of mind. It calls for research that can tease out the details of non-discrete orders of theory of mind. This can in turn be useful for research on the development of this cognitive capacity.

Appendix A

The Canteen Dilemma

Time left to complete this page: **0:51**

These instructions will also be shown on the following pages.

Instructions for the game:

This game is about trying to do the same as your colleague.

Every morning you arrive at work between 8:00 am and 9:10 am. You and your colleague will arrive by bus 10 minutes apart. Example: You arrive at **8:40 am**. Your colleague may arrive at **8:30 am**, or **8:50 am**.

Both of you like to meet in the canteen for a coffee. If you arrive before 9:00 am, you have time to go to the canteen, but you should only go if your colleague goes to the canteen as well. If you or your colleague arrive at 9:00 am or after, you should go straight to your offices.

At the beginning of each round you will know only your own arrival time. You will have to decide whether to go to the canteen or to the office. As a general rule, you will maximize your payoff by honestly choosing the option you think your colleague will also choose.

Payoff and penalties:

You start the game with $10.00 and will have to pay various amounts of penalties in each round, depending on how well you both do. Your challenge is to have as much money left as possible when the game ends, after which the remaining amount is paid out to you as a bonus. The game ends after 10 rounds or when you or your colleague has no money left.

- **Both go to canteen**
  If you guessed correctly that both of you went to the canteen before 9:00 am, you pay a **small** penalty proportional to how **uncertain** you were, e.g.:
  - **-$0.69** if you were very uncertain.
  - **-$0.29** if you were somewhat certain.
  - **-$0.01** if you were very certain.
- **Both go to office**
  If you guessed correctly that both of you went to your offices, no matter what time, your penalty is **doubled** and proportional to how **uncertain** you were, e.g.:
  - **-$1.39** if you were very uncertain.
  - **-$0.58** if you were somewhat certain.
  - **-$0.02** if you were very certain.
- **One goes to the canteen, the other to the office**
  If you guessed incorrectly and one of you went to the canteen while the other went to the office - or if any of you went to the canteen at 9:00 am or after, your penalty is **doubled** and proportional to how **certain** you were, e.g.:
  - **-$1.39** if you were very uncertain.
  - **-$2.77** if you were somewhat certain.
  - **-$9.21** if you were very certain.
- In summary, try to do your best doing the same as your colleague. As a general rule you will minimize your losses by giving an honest estimate of the chances of doing the same as your colleague

Next

Appendix B

Question 1: “The game is over. Do you think it was your fault it is over, your colleagues fault, or do you think it was because of some other reason?” with the possible answers being “Yes”, “No” and “Other reason”.

Question 2: “What strategy did you use while playing this game?”, where participants could answer in free text.

Question 3: “Imagine you could have agreed beforehand with your colleague about a point in time where it is safe to go to the canteen. What time would that be?”. The possible answers where: “I don’t know”, “There is no such time”, “8:00”, “8:10”, “8:20”, “8:30”, “8:40”, “8:50”, “9:00” and “9:10”.

Question 4: “Imagine you arrive at 8:00 am. Is it common knowledge between you and your colleague that it is safe to go to the canteen, that is, you both arrived before 9:00 am? “. The possible answers were: “Yes”, “No”, “I do not know”.

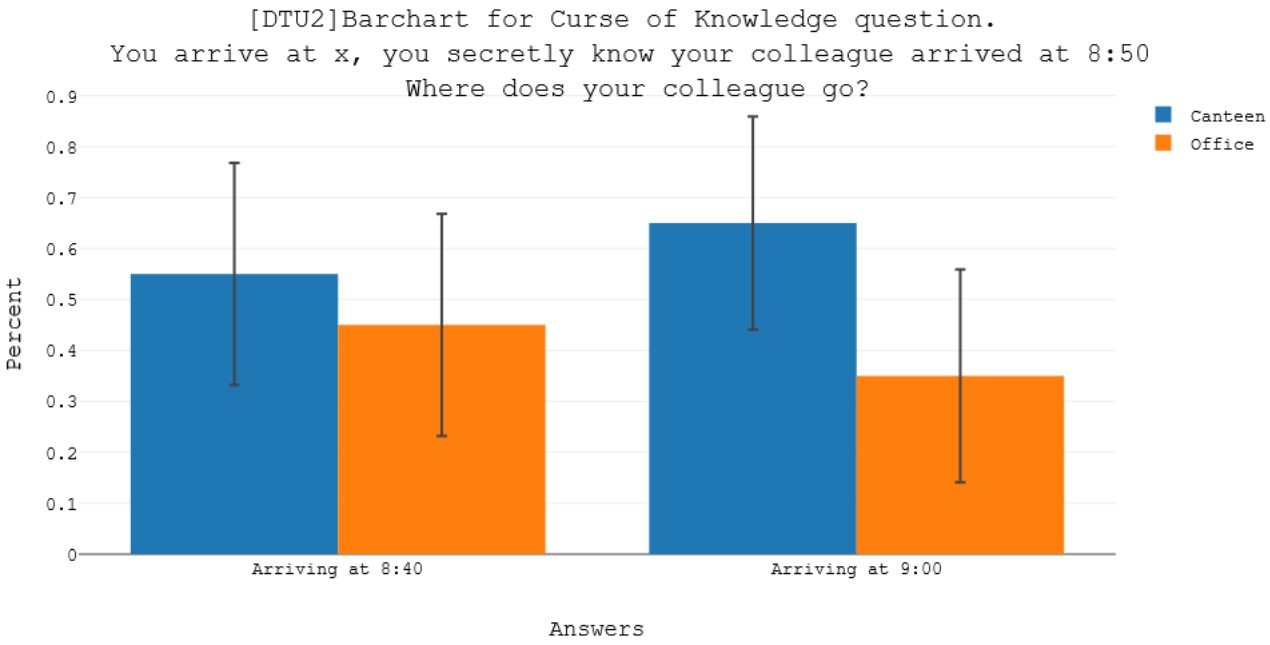
The DTU trials included the further questions:

Question 5: “Did you ever go to the canteen at an arrival time later than what was safe according to your previous answer? Why or why not?” Free text answer

Question 6: “Did you ever choose differently after seeing the same arrival time again at a later point in the game? Why or why not?" Free text answer

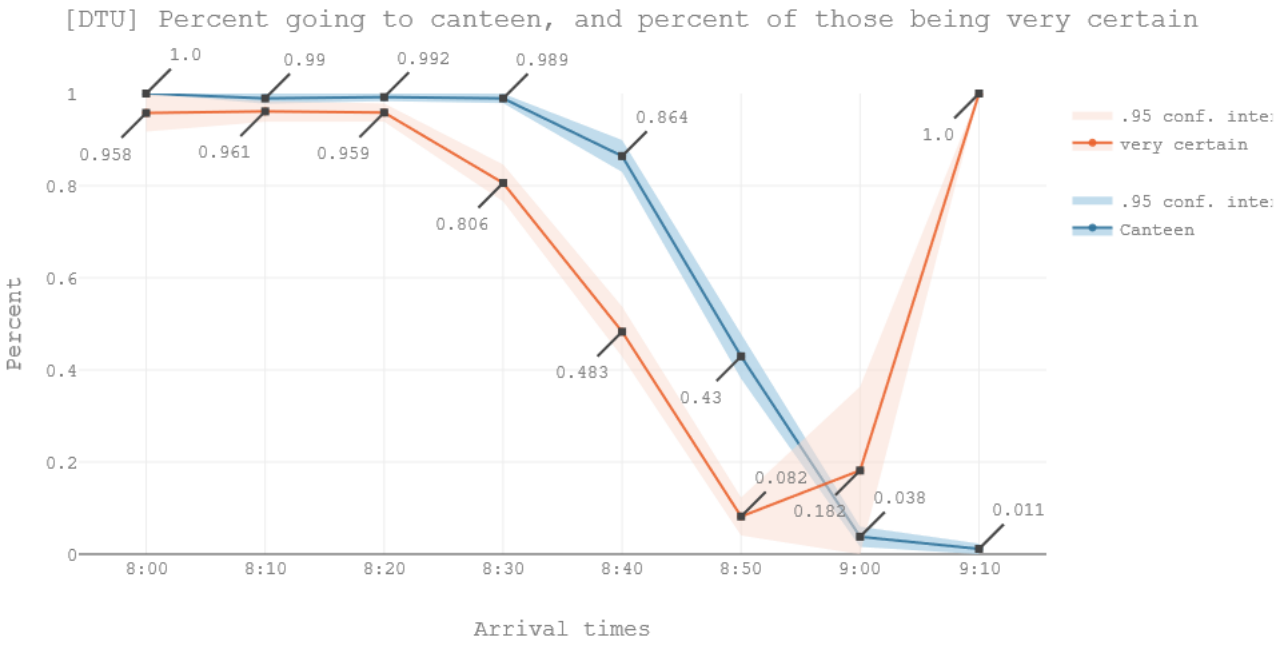
Question 7: “Imagine you arrived at 9:00//8:40 and you have been secretly informed that your colleague’s arrival time is 8:50. Where do you think your colleague will go?” Possible answers canteen/office. Half the participants got 8:40 and the other got 9:00

Appendix C

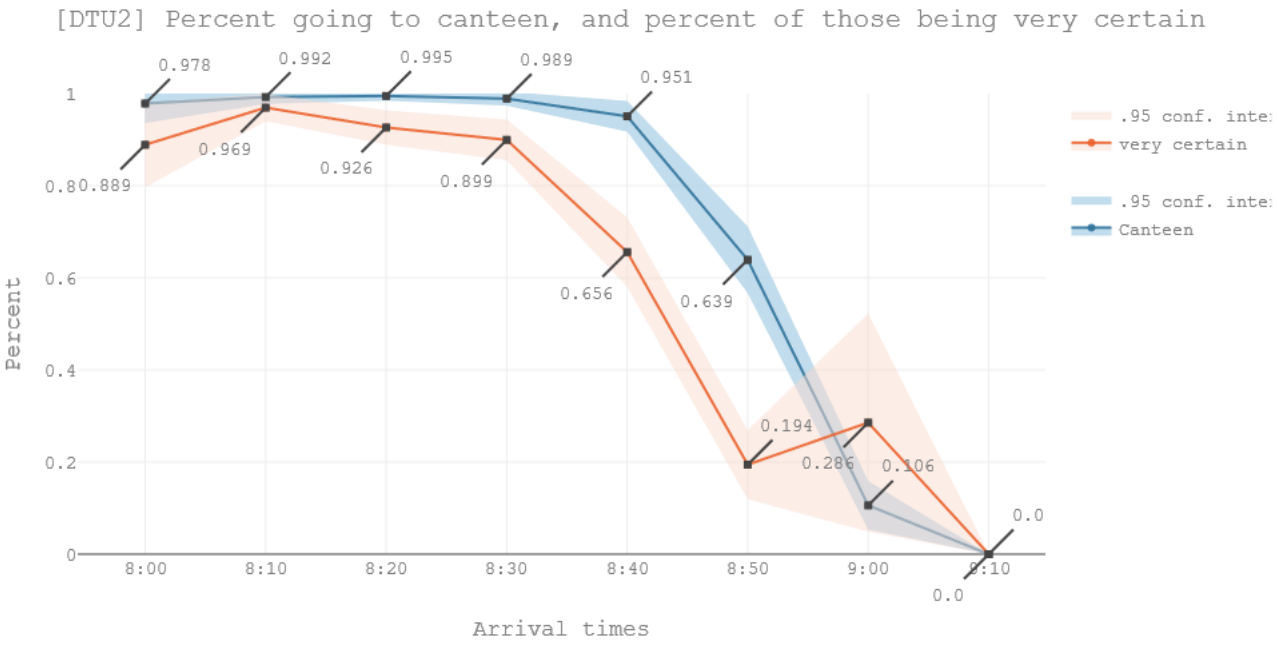


Grouped bar-chart for curse of knowledge question from second DTU trial.

Appendix C


[DTU] Percent going to canteen, and percent of those being very certain
Percent
Arrival times
8:00
8:10
8:20
8:30
8:40
8:50
9:00
9:10
0
0.2
0.4
0.6
0.8
1
1.0
0.99
0.992
0.989
0.864
0.958
0.961
0.959
0.806
0.483
0.43
0.082
0.182
0.038
0.011
1.0
very certain
Canteen

Appendix D



Appendix E

Appendix F

References

[1] Anderson, R. L. (2005). *Neo-Kantianism and the Roots of Anti-Psychologism*, British Journal for the History of Philosophy, 13:2, 287-323, DOI: 10.1080/09608780500069319

[2] Bacharach, M., & Stahl, D. O. (2000). *Variable-frame level-n theory.* Games and Economic Behavior, 32(2), 220–246.

[3] van Benthem, J. F. A. K. (2003). *Logic and the Dynamics of Information.* Minds and Machines 13: 503–519, Kluwer Academic Publishers

[4] van Benthem, J. F. A. K. (2007a). *Cognition as interaction.* In Proceedings symposium on cognitive foundations of interpretation (pp. 27–38). Amsterdam: KNAW.

[5] van Benthem, J. F. A. K., Gerbrandy, J., & Pacuit, E. (2007). *Merging frameworks for interaction: DEL and ETL.* In D. Samet (Ed.), Theoretical aspects of rationality and knowledge: Proceedings of the eleventh conference, TARK 2007 (pp. 72–81). Louvain-la-Neuve: Presses Universitaires de Louvain.*

[6] van Benthem, J. F. A. K., Hodges, H., & Hodges, W. (2007b). *Introduction.* Topoi, 26(1), 1–2. (Special issue on logic and psychology, edited by J.F.A.K. van Benthem, H. Hodges, and W. Hodges.).*

[7] van Benthem, J. F. A. K. (2008). *Logic and reasoning: Do the facts matter?* Studia Logica, 88, 67–84. (Special issue on logic and the new psychologism, edited by H. Leitgeb)

[8] van Benthem, J. F. A. K (2010). *Modal logic for open minds.* CSLI Publications.

[9] Benz, A., & van Rooij, R. (2007). O*ptimal assertions, and what they implicate. A uniform game theoretic approach.* Topoi, 26(1), 63–78 (Special issue on logic and psychology, edited by J.F.A.K. van Benthem, H. Hodges, and W. Hodges.)*

[10] Berinsky, A., Huber, G., & Lenz, G. (2012). *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. Political Analysis.* 20(3), 351-368. doi:10.1093/pan/mpr057

[11] Birch, S. A. J., Bloom, P. (2007). *The curse of knowledge in reasoning about false beliefs.* Psychol Sci. 2007 May; 18(5): 382–386. doi: 10.1111/j.1467-9280.2007.01909.x

[12] Buhrmester, Michael & Kwang, Tracy & Gosling, Samuel. (2011). *Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?.* Perspectives on Psychological Science. 6. 3-5. 10.1177/1745691610393980.

[13] Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). *An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use.* Perspectives on Psychological Science, 13(2), 149–154. https://doi.org/10.1177/1745691617706516

[14] Castelfranchi, C. (2004). *Reasons to believe: cognitive models of belief change.* Ms. ISTC-CNR, Roma. Invited lecture, Workshop Changing Minds, ILLC Amsterdam, October 2004. Extended version. Castelfranchi, Cristiano and Emiliano Lorini, The cognitive structure of surprise.Costa-Gomes, M., Weizsäcker, G., (2008). Stated beliefs and play in normal form games. Review of Economic Studies 75, 729–762.

[15] Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind. Child Development, 60, 1263–1277.

[16] Cheng P.W., Holyoak K.J, Nisbett R.E., Oliver L.M. (1986). *Pragmatic versus syntactic approaches to training deductive reasoning.* Cogn. Psychol. 18:293–328

[17] Chen, D.L., Schonger, M., Wickens, C., 2016. *oTree - An open-source platform for laboratory, online and field experiments.* Journal of Behavioral and Experimental Finance, vol 9: 88-97

[18] Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 362, 507–522.

[19] Crump M. J. C, McDonnell J. V., Gureckis T. M. (2013). *Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research.* PLoS ONE 8(3): e57410. https://doi.org/10.1371/journal.pone.0057410.

[20] Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). *Goal attribution without agency cues: the perception of 'pure reason' in infancy.* Cognition, 72, 237–267.*

[21] van Ditmarsch, H., van der Hoek, W., Kooi, B. (2008). *Dynamic Epistemic Logic.* Synthese Library, Springer Netherlands.

[22] van Ditmarsch H., Kooi B. (2015) *Consecutive Numbers.* In O*ne Hundred Prisoners and a Light Bulb.* Copernicus, Cham.

[23] Donkers, H. H. L. M., Uiterwijk, J. W. H. M., & van den Herik, H. J. (2005). *Selecting evaluation functions in opponent-model search.* Theoretical Computer Science, 349, 245–267.*

[24] Dunin-Keplicz, B., & Verbrugge, R. (2006). *Awareness as a vital ingredient of teamwork.* In P. Stone, & G. Weiss (Eds.), Proceedings of the fifth international joint conference on autonomous agents and multiagent systems (AAMAS'06) (pp. 1017–1024). New York: IEEE / ACM.*

[25] van Eijck, J., & Verbrugge, R. (Eds.) (2009). *Discourses on social software.* Texts in games and logic (Vol. 5). Amsterdam: Amsterdam University Press.

[26] Erb, Benjamin. (2016). *Artificial Intelligence & Theory of Mind.* 10.13140/RG.2.2.27105.71526.

[27] Fagin, R., & Halpern, J. (1988). *Belief, awareness, and limited reasoning.* Artificial Intelligence, 34, 39–76.*

[28] Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). Reasoning about knowledge, 2nd ed., 2003. Cambridge: MIT.

[29] Flobbe, L., Verbrugge, R., Hendriks, P.,&Krämer, I. (2008). *Children's application of theory of mind in reasoning and language.* Journal of Logic, Language and Information, 17, 417–442. (Special issue on formal models for real people, edited by M. Counihan.)*

[30] Frege, G. (1964 [1893]). *The Basic Laws of Arithmetic: Exposition of the System*, M. Furth (trans.), Berkeley, CA: University of California Press.

[31] Frege, G. (1897). *Logic*, reprinted in Frege [1997], pp. 227–250.

[32] Frege, G. (1997). *The Frege reader* (M. Beaney, editor), Blackwell, Oxford.

[33] Ghosh, S., Meijering, B., & Verbrugge, R. (2014). *Strategic reasoning: Building cognitive models from logical formulas*. Journal of Logic, Language and Information, 23(1), 1–29.

[34] Ghosh, S., Heifetz, A., & Verbrugge, R. (2015). Do players reason by forward induction in dynamic perfect information games? TARK.

[35] Ghosh, S., Meijering, B. & Verbrugge, R. (2018). *Studying strategies and types of players: experiments, logics and cognitive models.* Synthese (2018) 195: 4265. https://doi.org/10.1007/s11229-017-1338-7

[36] Gierasimczuk, N., Hendricks, V. F., Jongh, D. d. (2014). *Logic and Learning*. In Johan van Benthem on Logic and Information Dynamics, Baltag, Alexandru, Smets, Sonja (Eds.). Outstanding Contributions to Logic, Vol. 5. Dordrecht: Springer.

[37] Gigerenzer, G., Todd, P., & The ABC Research Group. (1999). *Simple Heuristics that Make us Smart*. New York: Oxford University Press.

[38] Griggs R.A., Cox J.R. (1982). *The elusive thematic-materials effect in Wason's selection task*. Br J Psychol 73:407–420

[39] Halpern, J. Y., & Moses, Y. (1990). K*nowledge and common knowledge in a distributed environment*. Journal of the ACM, 37, 549–587.*

[40] Harbers, M., Verbrugge, R., Sierra, C., & Debenham, J. (2008). *The examination of an information-based approach to trust*. In P. Noriega, & J. Padget (Eds.), Coordination, organizatiosn, institutions and norms in agent systems III. Lecture notes in computer science (Vol. 4870, pp. 71–82). Berlin: Springer.*

[41] Hedden, T., & Zhang, J. (2002). *What do you think I think you think? Strategic reasoning in matrix games*. Cognition, 85, 1–36.

[42] Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). *Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis*. Science, 317, 1360–1366.

[43] Horton, J.J., Rand, D.G. & Zeckhauser, R.J. (2011). *The online laboratory: conducting experiments in a real labor market*. Experimental Economics, Sep. 2014, Vol. 14: 399. https://doi.org/10.1007/s10683-011-9273-9

[44] Humphrey, N.K. (1980). *Nature's psychologists*. In Consciousness and the physical world (eds B. D. Josephson & V. S. Ramachandran), pp. 57–80. Oxford, UK: Pergamon Press.

[45] Isaac, A. M. C., Szymanik, J., & Verbrugge, R. (2014). *Logic and complexity in cognitive science*. In Johan van Benthem on Logic and Information Dynamics (pp. 787–824). Springer.*

[46] Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge: MIT.

[47] Keysar, B. & Lin, S. & J Barr, D. (2003). *Limits on theory of mind use in adults*. Cognition. 89. 25-41. 10.1016/S0010-0277(03)00064-7.

[48] van Lambalgen, M., & Counihan, M. (2008). *Formal models for real people*. Journal of Logic, Language and Information, 17, 385–389. (Special issue on formal models for real people, edited by M. Counihan).

[49] Leslie, A. (2000). How to acquire a 'representational theory of mind'. In D. Sperber & S. Davies (Eds.), Metarepresentation, Oxford: Oxford University Press.

[50] Lin, S., Keysar, B., Nicholas, E. (2010). *Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention.* Journal of Experimental Social Psychology Volume 46, Issue 3, May 2010, Pages 551-556.

[51] Liu, F. (2008). *Diversity of Agents and Their Interaction.* Springer Netherlands.

[52] McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *Proposal for the Dartmouth summer research project on artificial intelligence.* Technical report, Dartmouth College.

[53] Mason, Winter & Watts, Duncan. (2009). *Financial incentives and the performance of crowds.* SIGKDD Explorations. 11. 100-108. 10.1145/1600150.1600175.

[54] Maddy, P. (2012). *The philosophy of logic.* Bulletin of Symbolic Logic 18 (4):481-504.

[55] Meijering, B.,Maanen, L. v., Rijn,H. v., & Verbrugge, R. (2010). T*he facilitative effect of context on secondorder social reasoning.* In Proceedings of the 32nd annual meeting of the cognitive science society, (pp. 1423–1428). Philadelphia, PA, Cognitive Science Society.*

[56] Mol, L. (2004). Learning to reason about other people's minds. Technical report, Institute of Artificial Intelligence, University of Groningen, Groningen. Master's thesis.

[57] Pacuit, E., Parikh, R., & Cogan, E. (2006). *The logic of knowledge based obligation.* Synthese: Knowledge, Rationality and Action, 149, 57–87.*

[58] Palfrey, T., & Wang, S. (2009). *On eliciting beliefs in strategic games.* Journal of Economic Behavior & Organization, 71(2), 98-109.

[59] Parikh, R. (2003). *Levels of knowledge, games, and group action.* Research in Economics, 57, 267–281.

[60] Perner, J. (1988). *Higher-order beliefs and intentions in children's understanding of social interaction.* In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), Developing theories of mind (pp. 271–294). Cambridge: Cambridge University Press.

[61] Putnam, H. (1978). *There is at least one a priori truth.* Erkenntnis 13 (1978) 153-170.

[62] Quine, W. V. O (1951). *Two dogmas of empiricism.* Reprinted in his From a logical point of view, second ed., Harvard University Press, Cambridge, MA, 1980, pp. 20–46.

[63] Rand, David. (2011). *The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments.* Journal of theoretical biology. 299. 172-9. 10.1016/j.jtbi.2011.03.004.

[64] Rosenthal, R. (1981). *Games of perfect information, predatory pricing, and the chain store.* Journal of Economic Theory, 25, 92–100.*

[65] Seidenfeld, T., 1985. *Calibration, coherence, and scoring rules.* Philosophy of Science 52, 274–294.

[66] Stahl, D. O., & Wilson, P. W. (1995). *On players' models of other players: Theory and experimental evidence.* Games and Economic Behavior, 10, 218–254.

[67] Stenning K, van Lambalgen M. (2008).*Human reasoning and cognitive science.* MIT Press, Cambridge

[68] Stulp, F., & Verbrugge, R. (2002). *A knowledge-based algorithm for the internet protocol TCP*. Bulletin of Economic Research, 54(1), 69–94.*

[69] Sycara, K. & Lewis, M. (2004). *Integrating intelligent agents into human teams*. In E. Salas,& S. Fiore (Eds.), Team cognition: Understanding the factors that drive process and performance (pp. 203–232). Washington, DC: American Psychological Association. 133.

[70] Verbrugge, R., & Mol, L. (2008). *Learning to apply theory of mind*. Journal of Logic, Language and Information, 17, 489–511. (Special issue on formal models for real people, edited by M. Counihan.)

[71] Verbrugge R. (2009): *Logic and Social Cognition*. Journal of Philosophical Logic.

[72] Wason, P. C. (1966). *Reasoning*. In B. M. Foss (Ed.), New Horizons in Psychology I, (pp. 135–151). Harmondsworth: Penguin.

[73] Wason P.C., Shapiro D. (1971). *Natural and contrived experience in a reasoning problem*. Q J Exp Psychol 23:63–71

[74] Wason, P., & Shapiro, D. (1971). *Natural and contrived experience in a reasoning problem*. The Quarterly Journal of Experimental Psychology, 23(1), 63–71.

[75] Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

[76] Wimmer, H., & Perner, J. (1983). *Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception*. Cognition, 13, 103–128.

[77] Wooldridge, M. J. (2002). *An introduction to multiagent systems*. Chichester: Wiley.*

[78] http://www.glascherlab.org/social-decisionmaking/