Protocol for Canteen Dilemma

Robin Engelhardt, Thomas S. Nicolet

January 28, 2019

Contents

1 Background

The Canteen Dilemma experiment is meant to investigate what happens when people try to coordinate their actions without common knowledge. In the research literature, group G have *common knowledge* that A when everyone in group G knows A, and everyone in group G knows that everyone in group G knows A, and everyone in group G knows that everyone in group G knows that everyone in group G knows A etc. etc. In everyday use of the concept, common knowledge often refers to either something that everyone knows (called *mutual knowledge* in the research literature) or simply something that most people believe. This means there is a significant contrast between mutual knowledge and common knowledge.

The Canteen Dilemma is modeled after the Consecutive Numbers example in (Ditmarsch et al., 2008) which involves iterated reasoning about the reasoning of others. In this example, two agents, a (Anne) and b (Bill), face each other and can each see a number on the other's head. The numbers are consecutive numbers n and $n{+}1$ for a certain $n \in \mathbb{N}$, e.g. numbers 0 to 10, this face is common knowledge for Anne and Bill. Assume that Anne has the number 3 and Bill has 2. If either of them had a 0, the other would know their own number, which would have to be a 1. Now imagine that we ask them if it's common knowledge that they both have numbers higher than 0, e.g. that none of them know their number. Anne and Bill both know their number must be higher than 0, Since Anne see's a 2 and Bill sees a 3, so $E_{\{a,b\}}(\neg a_0 \wedge \neg b_0)$, where E denotes mutual knowledge. Now notice that since Anne can see a 2, she thinks it's possible she has 1 and Bill has 2: $\hat{K}_a(a_1 \wedge b_2)$, where $\hat{K}_a A$ means that a considers A possible, or that she does not know that A is false, e.g.$\hat{K}_a \varphi := \neg K_a \neg \varphi$. In that situation, Bill might not know his own number, since he sees a 1, but since he might have a 0 in that situation, Anne might know her number. So we have: $\neg E_{\{a,b\}} E_{\{a,b\}}(\neg a_0 \wedge \neg b_0)$. So while both Anne and Bill know that their numbers are higher than 0, it is not common knowledge. No instance of mutual knowledge can be common knowledge, regardless of how many iterations of mutual knowledge there is.

This is important in some cases, for example if we told Anne and Bill that they have to cooperate their actions in the following way. Assume that they get a large reward if both choose B and none of them have a 0, a smaller reward for both choosing A, and a large penalty for either choosing different options or choosing B when one of them has a 0. In our case of $(a_3 \wedge b_2)$,both would know they could choose B and get the largest reward. But since Anne sees a 2 on Bill, she thinks she might have a 1. If she has a 1, Bill would think he might have a 0, and he knows that Anne has to choose A if he has a 0. So since they don't have common knowledge about their numbers, the only safe strategy is choosing A, no matter what number they have. This is highly unintuitive and it is this aspect of social reasoning that the Canteen Dilemma is meant to capture.

In the Canteen Dilemma, respondents are told a story with a similar structure to the Consecutive Numbers game. Players are told that they and their colleague arrive at work every morning between 8:00 am and 9:10 am, and they always arrive 10 minutes apart. If they both arrive before 9:00 am, they have time to go to the canteen for a coffee. If either arrive at 9:00 am or after, they have to go to their offices. Players are rewarded if they both go to the canteen before 9:00 am, and get a smaller reward if they both go to their offices at any time. They get penalized if they either go to different places, or they go to the canteen at 9:00 am or later. Much like the consecutive numbers game, players only have common knowledge that they arrive 10 minutes apart, but no common knowledge about their arrival time.

The structure of the Canteen Dilemma can be found in coordination problems in everyday life when people communicate through text. In such cases, people might have mutual knowledge, but no proper common knowledge. This experiment is intended to investigate how people act and reason in such cases.

2 Research Questions

Common knowledge is an important theoretical concept in epistemic logic. It's not a clearly intuitive concept however, and it can be difficult for people to imagine cases where the difference between mutual knowledge and common knowledge makes a practical difference. Our experiment exhibits a structure where the lack of common knowledge makes a tangible difference. Our research questions and hypotheses relies on this fact and can be expressed like this:

1. The theoretical concept of common knowledge does not match our everyday notions of common knowledge. The infinite iteration about others beliefs does not occur in everyday situations and at most 2-3 iterations happen.

2. Most respondent will reason to approximately the same depth, which is 2-3 iterations of social reasoning. In other words, players will at most reason about others, reasoning about them.

3. Respondents will intentionally limit their reasoning depth because they find this reasoning depth sufficient for the task at hand. In other words, a limited reasoning depth is not due to lack of capacity in respondents but due to the belief that a certain reasoning depth is sufficient. Arriving early enough warrants that it is safe to go the canteen.

Our general research questions are: How much do respondents pick up on the importance of common knowledge? Assuming that respondents might act as if they had common knowledge, what is the optimal strategy for playing against sample of respondents?

3 Experimental Design and Setup

3.1 Instructions to player (requires formatting layout of screencaps)

The experimental setup is facilitated through Amazon's Mechanical Turk. After accepting the Amazon Mechanical Turk 'hit' all participants are asked to provide informed consent during the registration process. Upon being provided with a link to a UCPH-server at https://cibs.mef.sc.ku.dk and entering the experimental platform, participants wait until the choice room is available. A web page appears showing the following:

The Canteen Dilemma

Time left to complete this page: **24:50**

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 8:00am and 9:10am. You and your colleague will arrive by bus 10 minutes apart.
*Example: You arrive at **8:40am**. Your colleague can arrive at **8:30am**, or **8:50am**.*

Both of you like to meet in the canteen for a coffee. If you arrive before 9:00am, 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 9:00am 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.

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 9:00am, 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 9:00am 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

After seeing the instructions, the following web page appears.

Round 1

It is Monday morning and you arrive at your workplace at 8.10 am.

Where will you go?

Canteen Office

After making their choice, they have to give an estimate of their certainty that the other player made the same choice.

How certain are you that your colleague has made the same choice as you?

- ○ Very uncertain
- ○ Slightly certain
- ○ Somewhat certain
- ○ Quite certain
- ○ Very certain

Next

Each round, after having made their choice and estimated their certainty, respondents see the following result page:

Results (after round 4)

Round	You went to the	at	Your colleague went to the	at	How certain you were that colleague did the same as you	**Penalty**
1	office	8.10	office	8.20	very certain	**-$0.02**
2	office	9.00	office	8.50	somewhat certain	**-$0.58**
3	canteen	8.20	canteen	8.10	quite certain	**-$0.13**
4	office	8.50	canteen	8.40	somewhat certain	**-$2.77**

You went to the office while your colleague did not.

You lost **-$2.77** this round. Your total losses are **-$3.50**. You still have **$6.50**.

Next

The game stops if either player loses their initial $10 bonus

3.2 Questions

The experiment involves two key questions and five supplementary ones. Each round involves these two questions:

- It is Monday morning, and you have arrived at your workplace at 8:10 am. Where will you go? (Either the canteen or the office).
- How certain are you that your colleague has made the same choice as you? (Very uncertain, slightly certain, somewhat certain, quite certain, very certain).

If the game is lost because either respondent has lost their initial bonus, they are given the question:

- 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? (My fault, my colleagues fault, other reason)

When the game is either over after 10 rounds or is lost, the respondents receive the following questions:

- What strategy did you use while playing the game? (Free text)
- Assume that you prior to the game could have decided a cut-off point with your colleage which you would both stick to while giving a *very certain* estimation of making the same choice. This means .meaning that you will only go to the office if you arrive at this time or later. At what time would this cut-off point be? (Do not know, none of these times, 8:00, 8:10, 8:20, 8:30, 8:40, 8:50, 9:00, 9:10).
- Question about everyday notion of common knowledge
- Question about theoretical notion of common knowledge

3.3 Study Type

The experiment will be done online on the Amazon Mechanical Turk Platform. Participants will be from all over the world. The software platform is otree 2.1.7

3.4 Inclusion and exclusion rules on Mturk

No player is allowed to play the game more than once which we will secure by giving each respondent a permanent qualification which excludes them from further participation.

4 Treatments / Conditions

We only employ one treatment so far, the Canteen Dilemma with a logarithmic scoring rule.

5 Payoff Structure

5.1 General structure:

Respondents receive a participation fee of X plus and an initial $10 bonus. Every round they will pay a penalty which is taken out of their bonus depending on the choices they make and how certain they are that their colleague has made the same choice. Payoffs range from $0.01 to $9.21. In order to get respondents to answer honestly, logarithmic scoring is used as a proper scoring rule.

If both players go to the office, their penalty is $\ln(player\_certainty) \cdot 2$, if they borh go to the canteen it is $\ln(player\_certainty)$ and if they go to different places it is $\ln(1 - player\_certainty) \cdot 2$. We expect that many will lose their entire bonus, while other's might keep some of their bonus if their co-player loses all their bonuses early and ends the game.

(Probably skip next section of examples of scores)

5.2 Payoff for both going to the office:

Calculated by $\ln(player\_certainty) \cdot 2$

- Very uncertain (50%) = - $1.39
- Slightly uncertain (62.5%) = -$0.94
- Somewhat certain (75%) = -$0.57
- Quite certain (87.5%) = -$0.27
- Very certain (99%) = -$0.02

5.3 Payoff for both going to the canteen

Calculated by $\ln(player\_certainty)$

- Very uncertain (50%) = - $0.69
- Slightly uncertain (62.5%) = -$0.47
- Somewhat certain (75%) = -$0.29
- Quite certain (87.5%) = -$0.13
- Very certain (99%) = -$0.01

5.4 Payoff for both going to the canteen

Calculated by $\ln(1 - player\_certainty) \cdot 2$

- Very uncertain (50%) = - \$1.39
- Slightly uncertain (62.5%) = -\$1.96
- Somewhat certain (75%) = -\$2.77
- Quite certain (87.5%) = -\$4.16
- Very certain (99%) = -\$9.21

6 Theoretical Approach and Methods (expectations, probably reduce this, maybe brief focus on Theory of Mind instead)

We strongly expect that respondents always choose to go to the office when they arrive at 9:00 am or 9:10 am. We can say that for any player a, if a arrives at 9:00 am or 9:10 am, we can write $K_a(late)$, where *late* means that it's too late to go to the canteen because either player arrived at 9:00 am or after. We have a strong assumption that $K_a(late)$ entails that a will go to the office. It's more complicated what $K_a(\neg late)$ entails for a's decision.

Imagine a arrives at 8:50 am. She then knows b arrived at 8:40 am or 9:00 am, i.e. a doesn't know if it's too late to go to the canteen or not: $\neg K_a(\neg late) \wedge \neg K_a(late)$. We can read $\neg K_a(\neg late)$ as expressing that a considers it possible that it's too late to go to the canteen, written as $\hat{K}_a(late)$, given this definition$\hat{K}_a\varphi := \neg K_a \neg\varphi$.

Continuing this line of reasoning, assume a arrives at 8:40 am. In this case, a knows b arrived at 8:40 or 8:50, which means he knows it's not too late to go to the canteen: $K_a(\neg late)$. But a doesn't know if b knows this, since she considers it possible that b arrived at 8:50 am, in which case b would not know if it is too late to go to the canteen, so we have $\hat{K_a}\hat{K_b}(late)$ or otherwise expressed as $K_a \neg E_{\{a,b\}}(\neg late)$. In this case, a knows it's not too late to go to the canteen, but he does not know if b knows this.

Similar reasoning applies for a arriving at 8:30 am. In this case we have $K_a(\neg late)$ and given that b can arrive at 8:40 am at the latest, we have $K_b(\neg late)$ and a knowing this epistemic fact: $K_a K_b(\neg late)$. This implies $E_{\{a,b\}}(\neg late)$. Since a knows that both a and b knows it's not too late to go to the canteen, we have $K_a E_{\{a,b\}}(\neg late)$. But if b arrives at 8:40, it holds that everyone knows it's not too late to go to the canteen, but that not everyone knows this very fact! I.e. $E_{\{a,b\}}(\neg late) \wedge \neg E_{\{a,b\}} E_{\{a,b\}}(\neg late)$.

The interesting aspect is what knowledge will be sufficient for a player to go to the canteen. If player a arrives at 8:50, she will consider it possible that it is too late to go to the canteen, i.e. $\hat{K}_a(late)$. We might have the assumption that$\hat{K}_a(late)$ entails that a will go to the office. But if a arrives at 8:40 am, she will consider it possible that b considers it possible that it's too late: $\hat{K}_a\hat{K}_b(late)$. If b reasons like a, then a would have to go to the office when arriving at 8:40 as well. Since the backwards deduction doesn't stop there, the same holds for any previous time, meaning there might be no knowledge sufficient for going to the canteen.

This line of reasoning is highly unintuitive, meaning we expect respondents to be limited to 1-3 layers of social reasoning, i.e. reasoning about other's reasoning. Secondly, even if a respondent know that there is no have common knowledge about arriving before 9:am, they know that this fact is not common knowledge among the players, meaning they might go to the canteen before 9:00 am as well.

Note however that if both players go to the office with a 'very certain' estimate every time, they will be certain to get a bonus of $9.80. If players arrive before 9:00 am at all times and go to the canteen with a 'very certainty' estimate, they get a bonus of $9.90, only $0.10 more than the safe *office only* strategy. And any strategy involving going to the canteen, e.g. at any time before 9:00 am, is likely to be given times at 8:50 am and 9:00 am, where the players will not go to the same place. Instead of going to the canteen at 8:50 am or earlier, players might move a supposed 'cut off' point to an earlier time. But this only exchanges instances of successfully going to the canteen with instances of successfully going to the office, while still experiencing the risk of making different choices.

The only strong assumption we have is that for any agent m, $K_m(late)$ entails that m will go to the office. We also know that any strategy which involves going to the canteen at some time will include a risk of making a different choice than the other player.

7 Existing knowledge

Most of current research literature on social reasoning suggest the cognitive human capacity is limited at 2 iterations, i.e. I reason about you, reasoning about me, while few reports find higher levels of reasoning [**?**]

8 (AMT Type)

9 Ethical Considerations

9.1 Approval by the Institutional Review Board

We provide participants with a consent page before the experiment starts and participants need to check a box in order to proceed. If not, they are not allowed to participate in the experiment.

9.2 Anonymity and data management

We will follow the faculty's guidelines in accordance with the Danish Data Protection Act, cf. requirements from the Danish Data Protection Agency and the faculty's Institutional Review Board.

Data will be stored in a separate database at https://cibs.mef.sc.ku.dk, where the experiment will take place. No data will be stored in the cloud.

10 Additional Considerations

10.1 Time Plan

10.2 Total Cost and Financing

References

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

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