Testing and Training Theory of Mind for Hybrid Human-Agent Environments

ICAART, Valletta, Malta, February 22-24, 2020

Rineke Verbrugge, Bernoulli Institute
Department of Artificial Intelligence

university of groningen

Putting yourself in someone else's shoes

"I know you think you understand
what you thought I said
but I'm not sure you realize that
what you heard is not what I meant."
(Alan Greenspan)

For artificial agents to effectively interact with people in *Hybrid Intelligence* – whether it be competition, teamwork or negotiation – they will need to put themselves in the shoes of these people.
This includes modeling people's mental models of others.

What is Theory of Mind?

The ability to reason about *mental states* of others

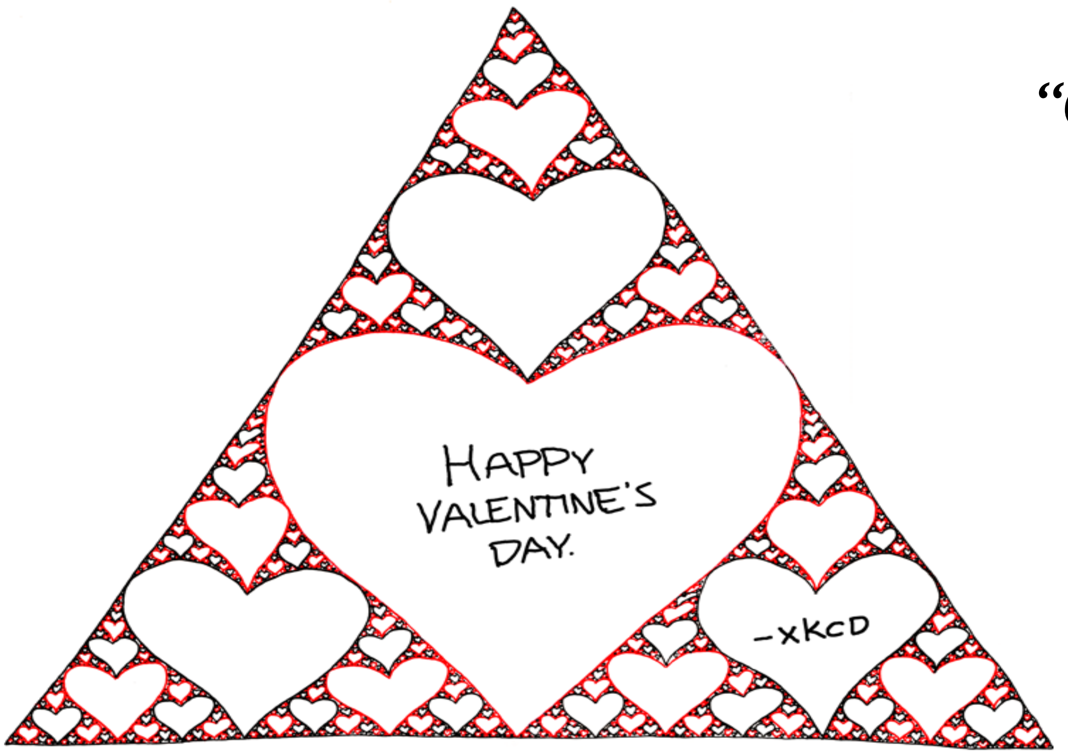
This may concern their beliefs, thoughts, knowledge, intentions

People use it to explain, predict and manipulate behavior of others

People apply it recursively:
higher-order theory of mind

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *BBS, 1*,515-526
Dennett, D. *Intentional Systems*. Cambridge (MA), MIT Press, 1989

Orders of theory of mind



- *0-order theory of mind:*
  "Chris was the one who sent you the anonymous Valentine's card"
  (abbreviation: p)

- *1st-order theory of mind:*
  *"You **know** that* p*"*: $K_Y\ p$

- *2nd-order theory of mind:*
  *"Does Chris **know** that you **know** that* p*?"*
  $K_C\ K_Y\ p$

Luc Steels' birthday puzzle

The following is common knowledge:

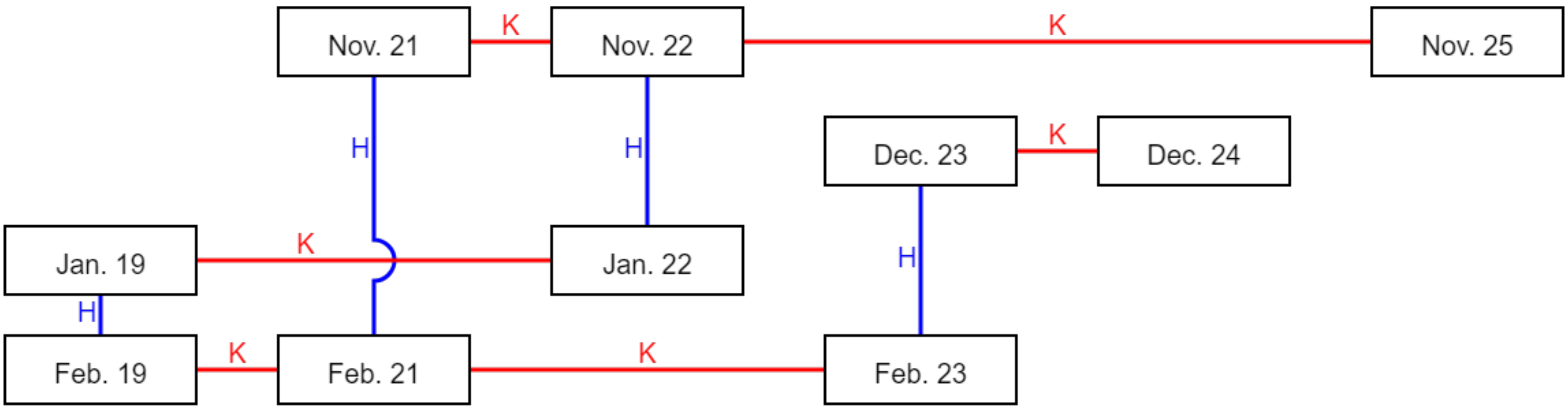
- Luc Steels' birthday is one of the following 10 dates:
  November 21, 22 or 25; or
  December 23 or 24; or
  January 19 or 22; or
  February 19, 21 or 23
- Kim and Harmen are perfect logicians and they always speak the truth (just like Rineke)
- Kim knows the month of Luc Steels' birthday (possibilities: Nov, Dec, Jan, Feb)
- Harmen knows the day of Luc Steels' birthday (possibilities: 19, 21, 22, 23, 24, 25)

The following dialogue takes place:

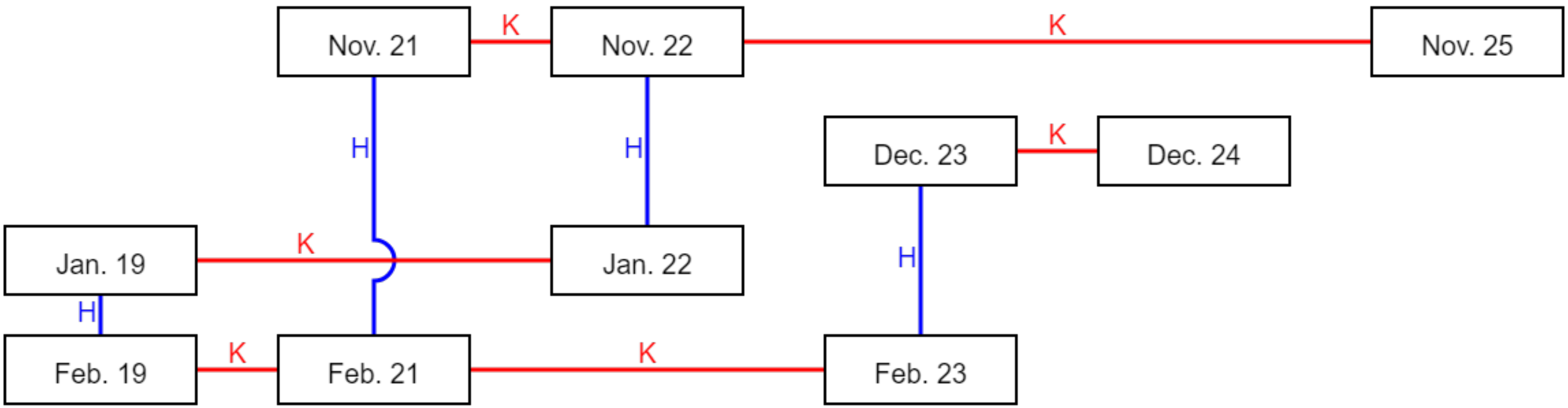
1. *Kim:* Harmen, I know that you do not know Luc Steels' birthday
2. *Harmen:* Now I do know Luc Steels' birthday
3. *Kim:* Now I know it, too!

When is Luc Steels' birthday?

Just before the dialogue

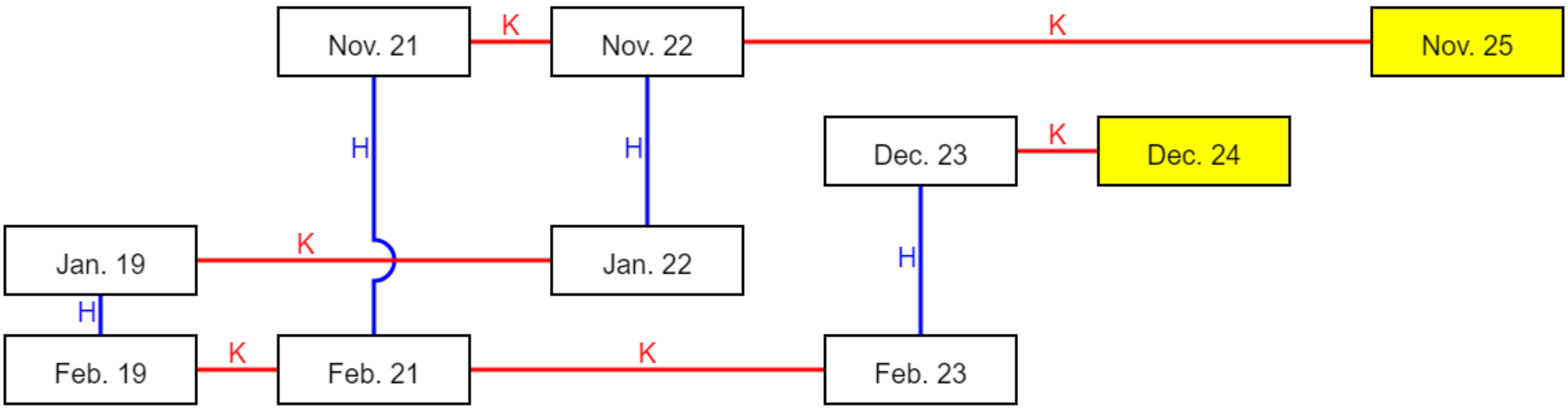


Just before the dialogue



Kim: Harmen, I know that you do not know Luc Steels' birthday

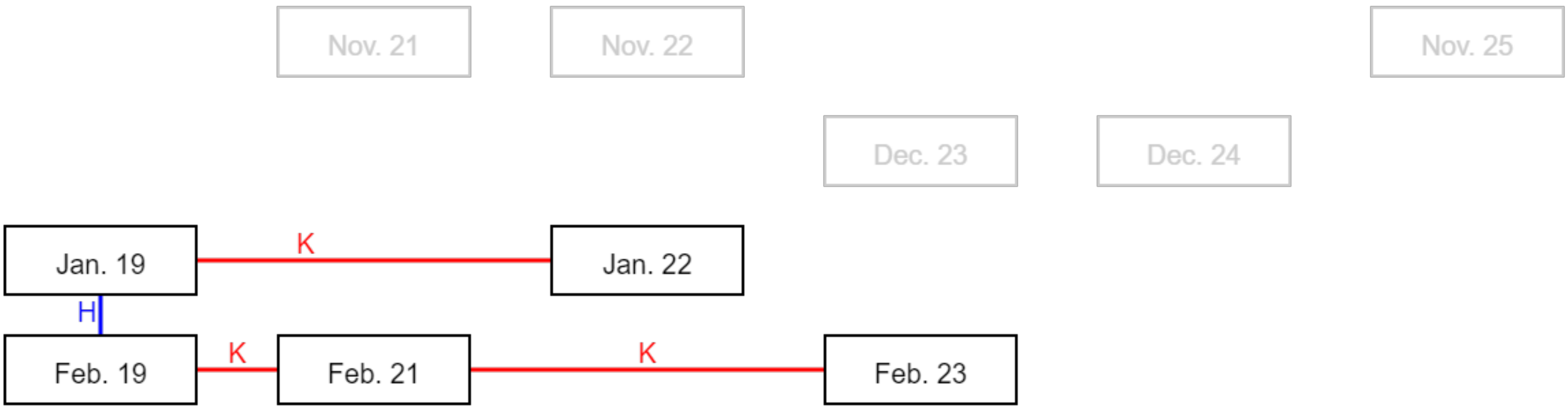
Just before the dialogue



Kim: Harmen, I know that you do <u>not</u> know Luc Steels' birthday

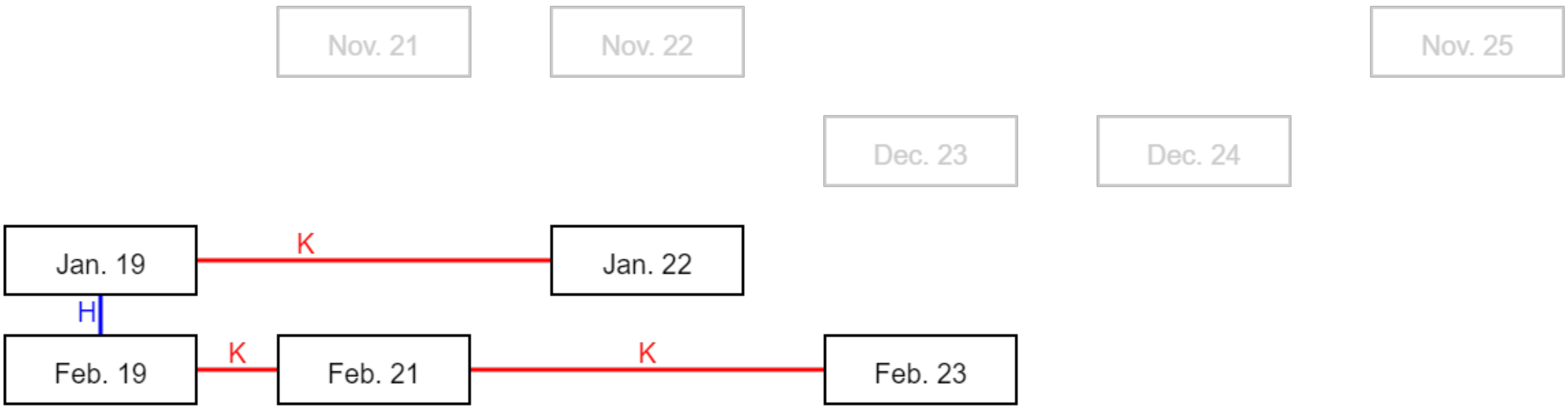
After the first announcement

Kim: Harmen, I know that you do not know Luc Steels' birthday



After the first announcement

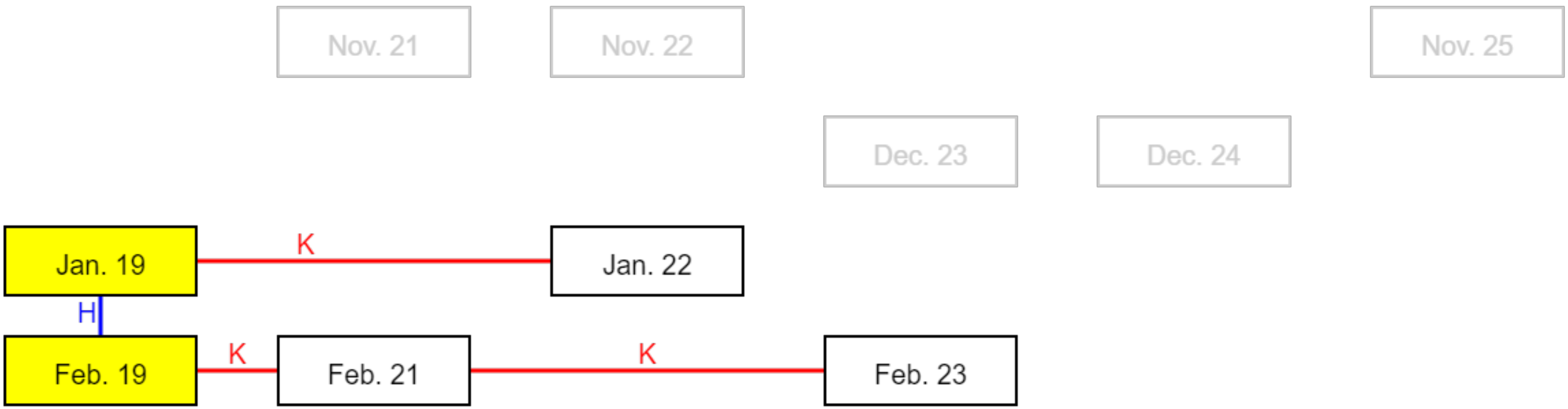
Kim: Harmen, I know that you do not know Luc Steels' birthday



Harmen: Now I do know Luc Steels' birthday

After the first announcement

Kim: Harmen, I know that you do not know Luc Steels' birthday

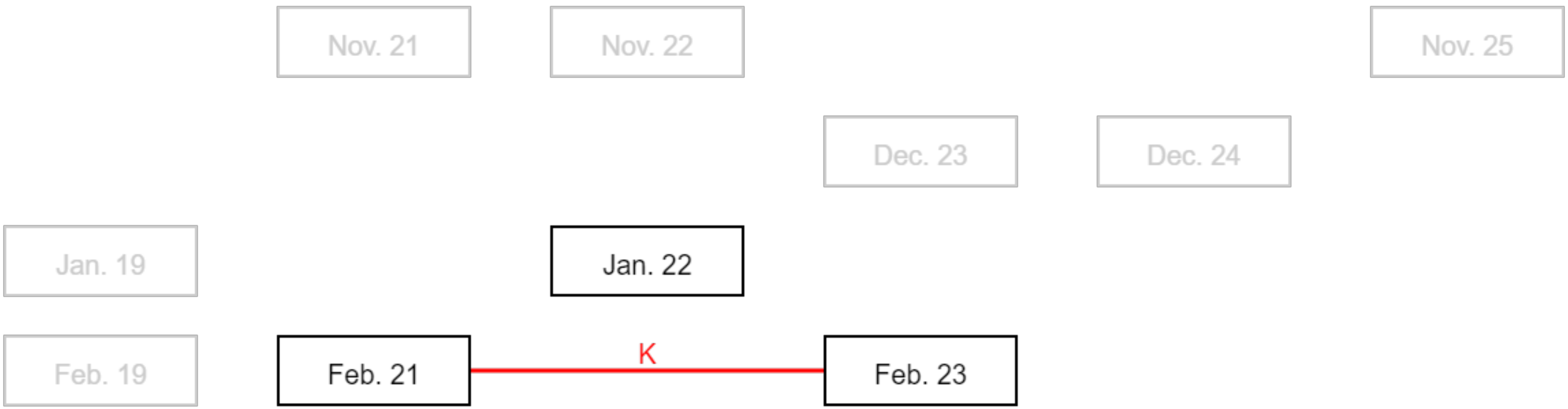


Harmen: Now I do know Luc Steels' birthday

After the first two announcements

Kim: Harmen, I know that you do not know Luc Steels' birthday

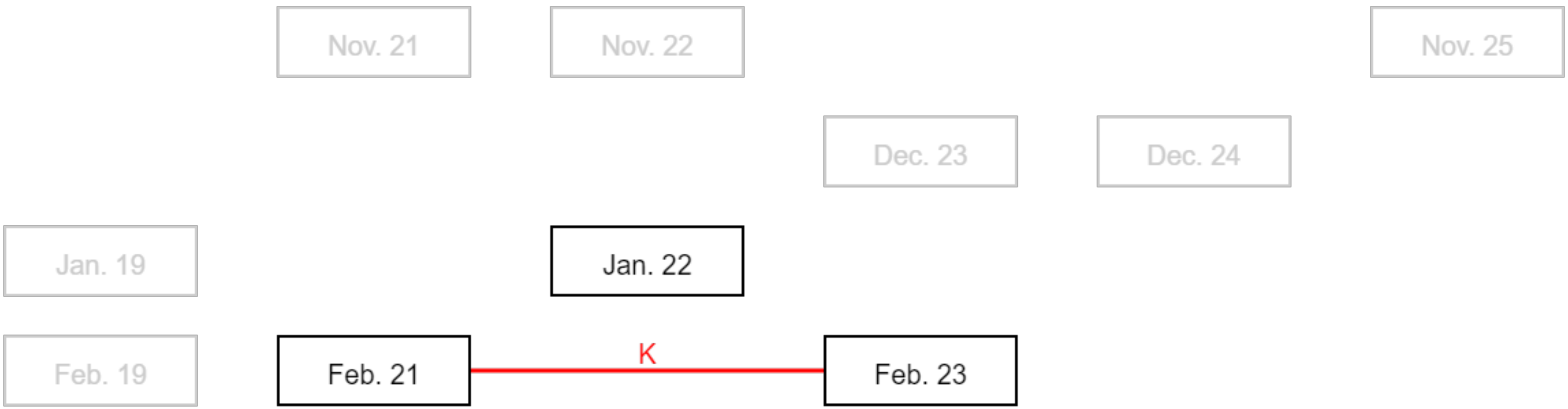
Harmen: Now I do know Luc Steels' birthday



After the first two announcements

Kim: Harmen, I know that you do not know Luc Steels' birthday

Harmen: Now I do know Luc Steels' birthday

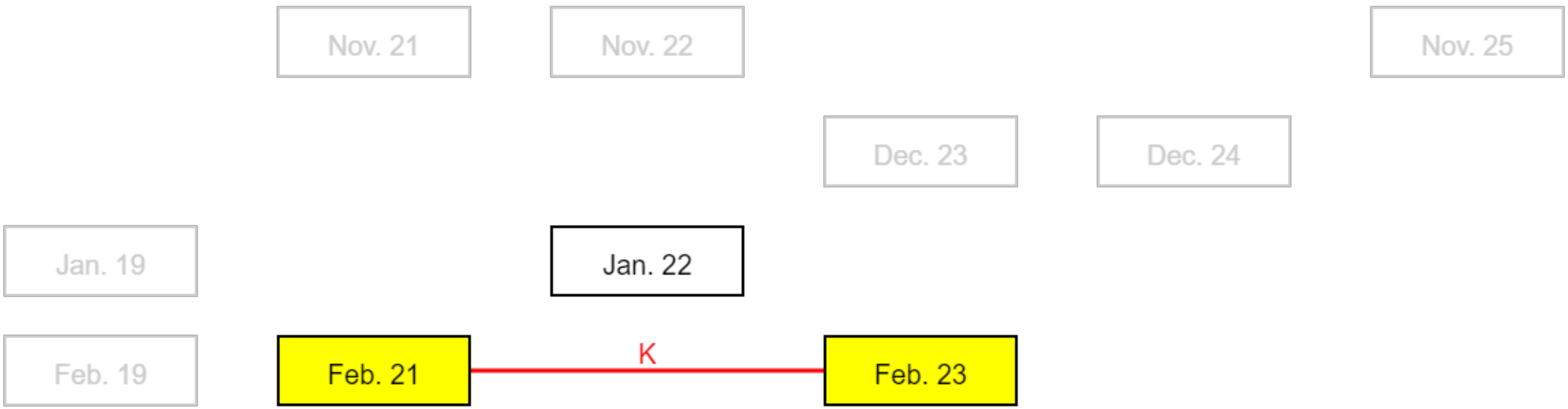


Kim: Now I know it, too!

After the first two announcements

Kim: Harmen, I know that you do not know Luc Steels' birthday

Harmen: Now I do know Luc Steels' birthday



Kim: Now I know it, too!

After all three announcements

Kim: Harmen, I know that you do not know Luc Steels' birthday

Harmen: Now I do know Luc Steels' birthday

Kim: Now I know it, too!

	Nov. 21	Nov. 22			Nov. 25
			Dec. 23	Dec. 24	
Jan. 19		**Jan. 22**			
Feb. 19	Feb. 21		Feb. 23		

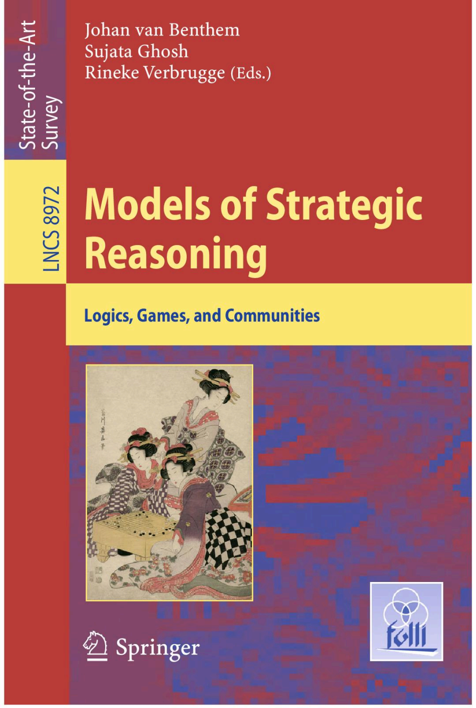
-H. v. Ditmarsch, J. Ruan & R. Verbrugge, Sum and product in dynamic epistemic logic. *Journal of Logic and Computation*, *18*(4) (2007) , 563-588.

Overview remainder of the talk

- **I. The Marble Drop game**
  - Training people to do better in the game by using 2nd-order ToM
  - Using eye movements and reaction times to estimate adults' reasoning strategies

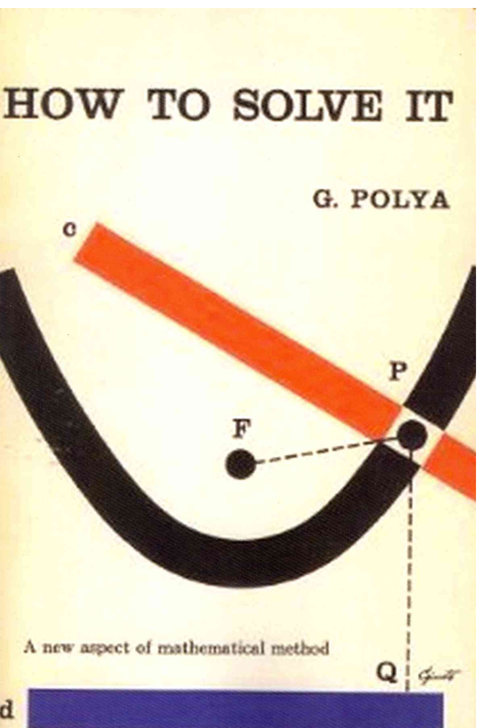
- **II. The Mod game**
  - Using random effects Bayesian model selection to fit participants' choices when playing against different software agents
  - Using software agents to invite them to play better

- **III. The Colored Trails game**
  - Using software agents to estimate a student's theory of mind *and* to entice them to use second-order ToM



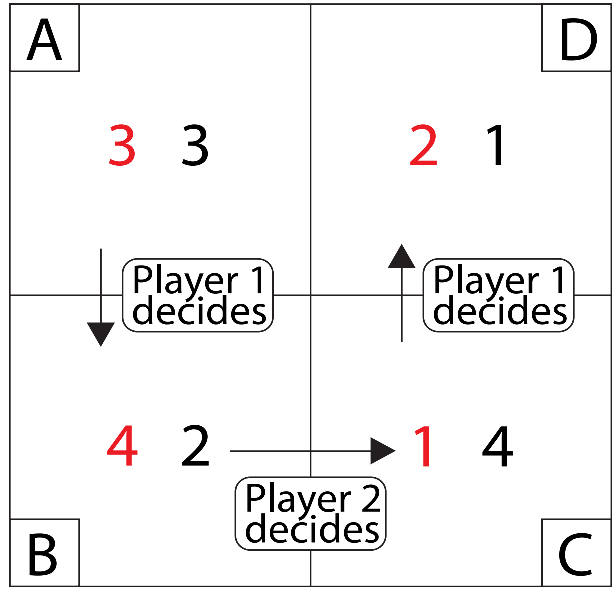
Strategic reasoning & Reasoning strategies

In game theory, a player's *strategy* is a partial function from the set of event histories at each stage to their set of actions in their turn to move. Agents choose a strategy for maximal gain.



In cognitive science, *strategy* is used more broadly, as in Polya's problem solving strategies (understanding the problem, developing a plan, performing the plan, and looking back). Cognitive scientists construct fine-grained theories about human reasoning strategies from experiments with human participants.

Applying second-order theory of mind in turn-taking games was thought to be notoriously difficult



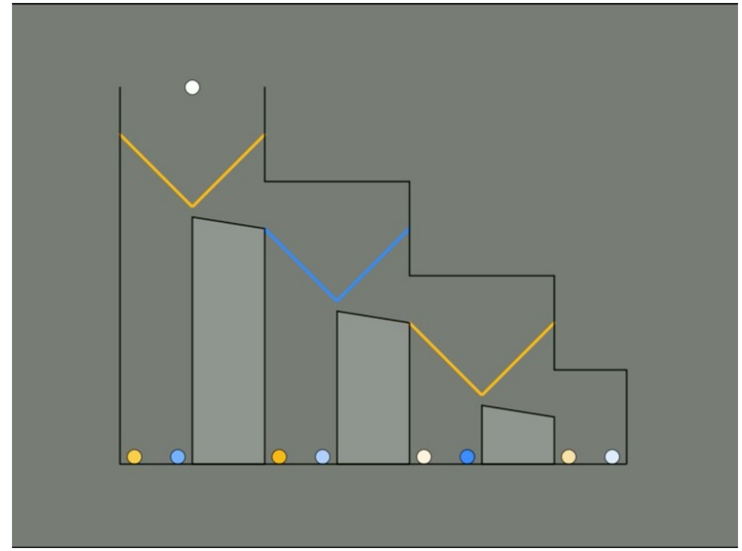
In the Matrix Game, Player 1 has to use 2nd-order ToM: "What does Player 2 *think* I *intend* to choose if the game gets to C? So, what to choose in A: Stay or move to B?"

In Hedden & Zhang's experiment, students who knowingly played against a *rational* computer opponent got only about 50% correct (e.g. move to B)

Hedden, T., & Zhang, J., What do you think I think you think? *Cognition* **85** (2002) 1-36

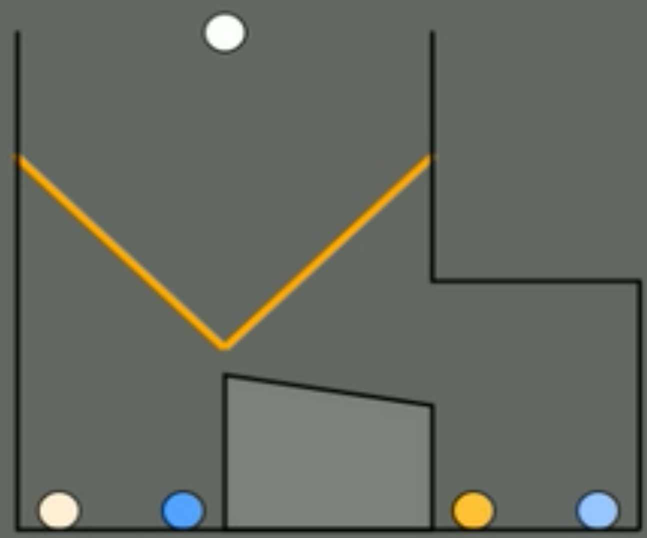
I. The Marble Drop game

We designed the game Marble Drop:

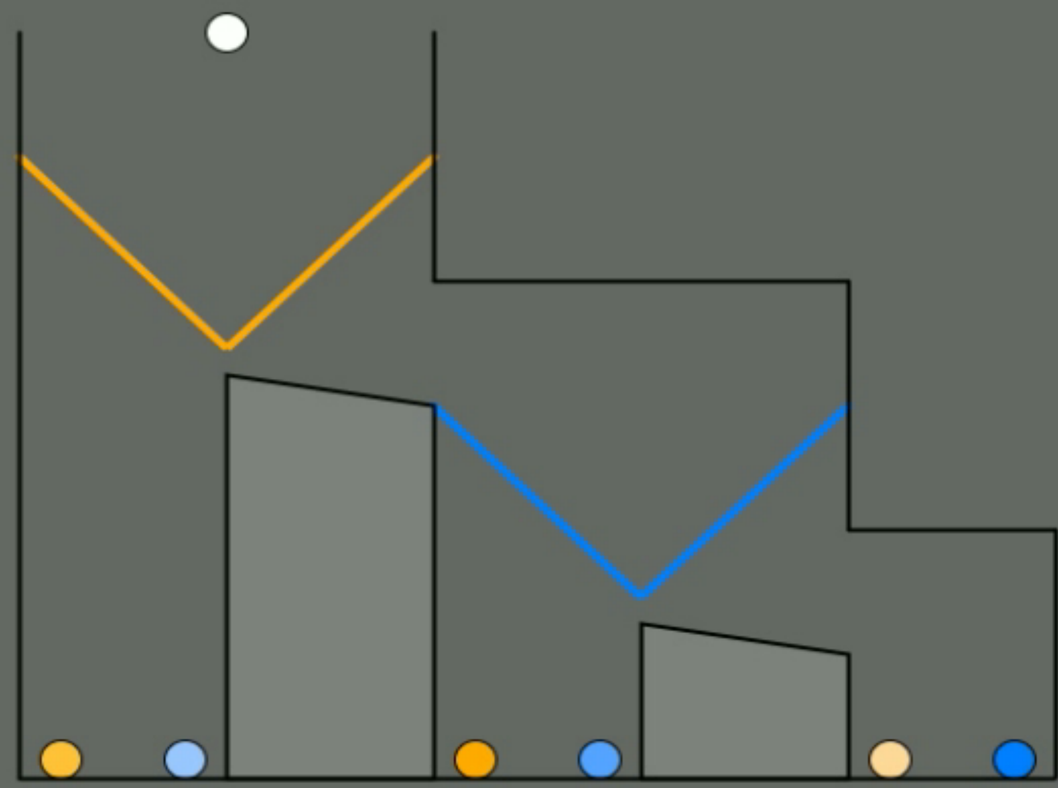


- A turn-taking game between
  the participant (orange) and a computer player (blue)
- A white marble drops down. Players control the course of the marble by opening the left or right trapdoor of their color
- The participant wants the marble to drop into a bin in which the left marble is as dark orange as possible
- The computer wants the marble to drop into a bin in which the right marble is as dark blue as possible

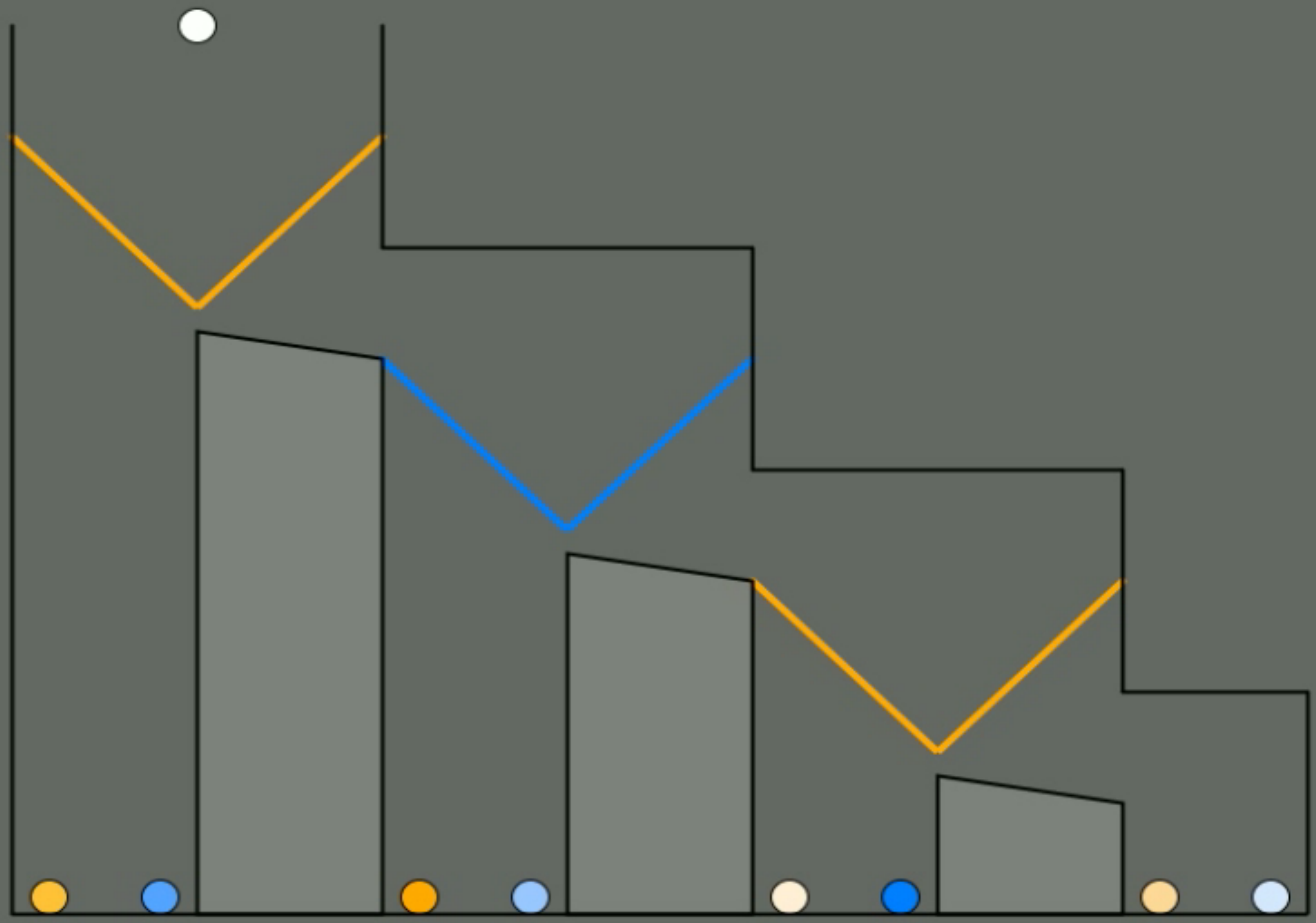
Zero-order Marble Drop game



First-order Marble Drop



Second-order Marble Drop

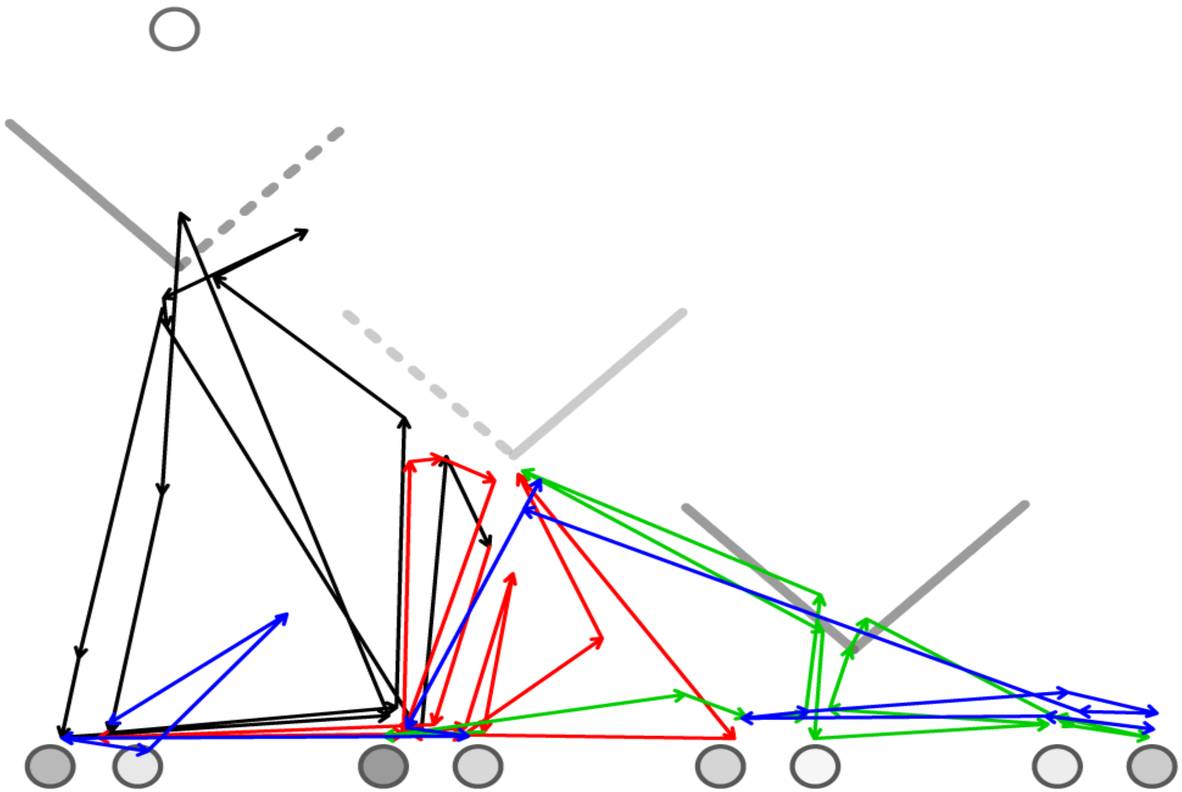


Results for participants playing Marble Drop with stepwise training

- Participants played 4 ToM_0, then 4 ToM_1, then 8 ToM_2 training games
- Participants then played 2 x 32 test games.
- They were asked to make their choice at the first set of orange trapdoors
- Test games were constructed to be diagnostic for ToM_2
- Games for which the optimal choice is 'left' and those for which it is 'right' were balanced
- Result: by the end of test block 2, 94% of participants' choices were optimal, corresponding to applying ToM_2

Verbrugge, R., Meijering, B., Wierda, S., van Rijn, H., & Taatgen, N. (2018). Stepwise training supports strategic second-order theory of mind in turn-taking games. *Judgment and Decision Making*, 13(1), 79-98

But how do the participants actually reason? Probably not by backward induction!

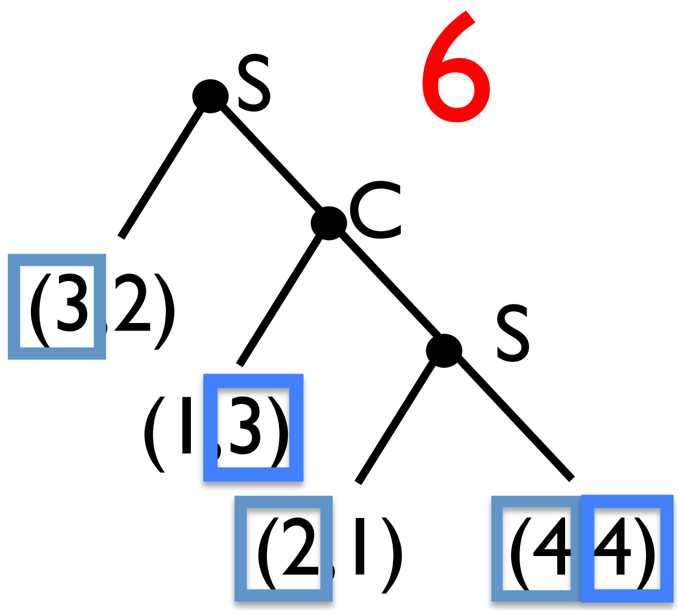


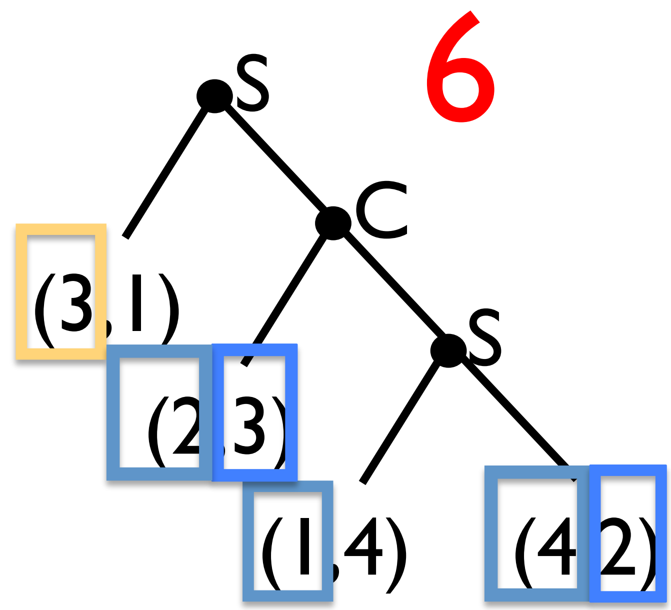
Eye movements of a typical participant
1) Black: fixations 1-15; 2) Red:, fixations 16 – 30;
3) Green: fixations 31 – 45; 4) Blue: fixations 46 – 61

B. Meijering, H. van Rijn, N. Taatgen and R. Verbrugge, What eye movements can tell about theory of mind in a strategic game, *PloS ONE 7 (9)*, 2012, e45961.

Backward induction

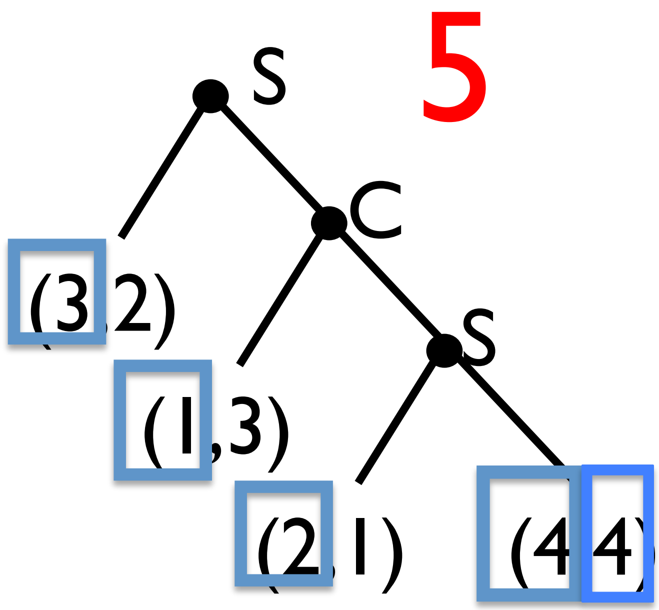
One step: to attend to a pay-off for comparison.
For every pay-off structure, 6 steps are needed.

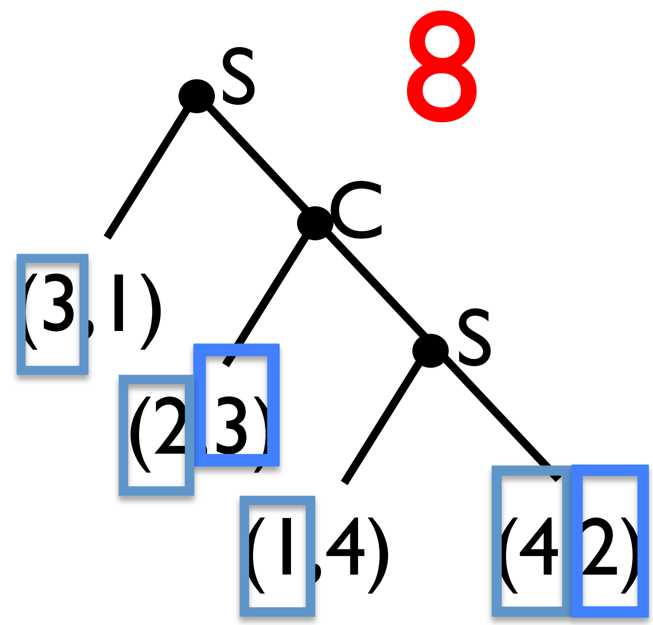




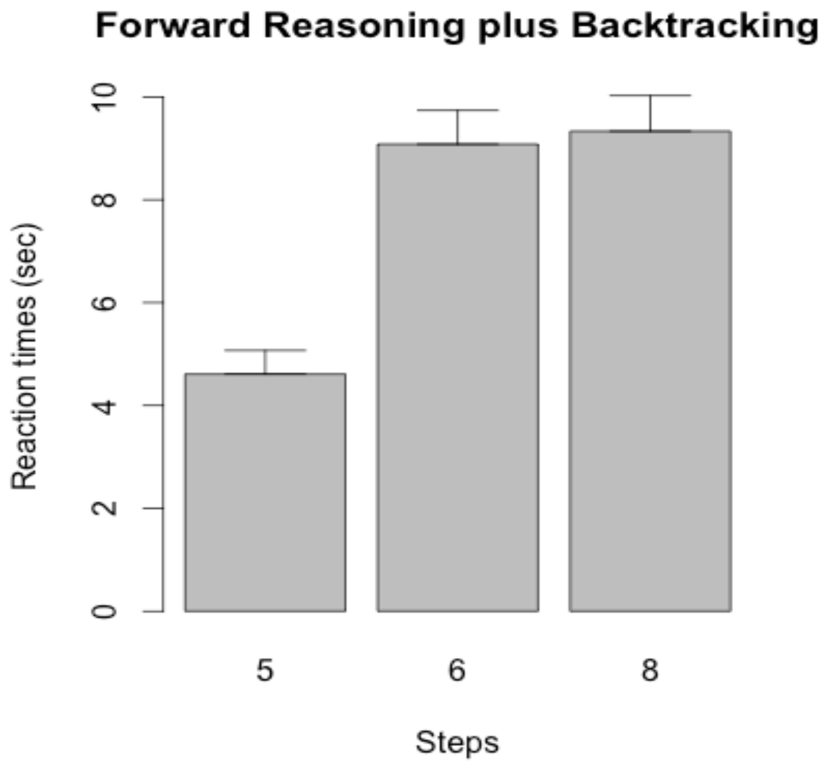
Forward-reasoning plus backtracking

Reason forward to see where is your highest payoff.
Backtrack to see whether it is attainable.
If last leaf has (4,4), you only need to check 5 pay-offs from the root down.
For other pay-off structures, 6 or 8 steps are needed





Reaction times for 5, 6 and 8 step games



Could it be that in many real-life strategic interactions, forward reasoning plus backtracking is faster than backward induction?

Or, maybe, people tend to reason in a forward way, as in commonsense reasoning about cause and effect.

G. Bergwerff, B. Meijering , Jakub Szymanik, Rineke Verbrugge and Stefan Wierda, Computational and algorithmic models of strategies in turn-based games, *Cognitive Science 2014*

II. The Mod game

-Repeated one-shot game
-Software agents entice people to use higher-order ToM
-Random effects Bayesian model selection for estimation of distribution of strategies

Kim Veltman, Harmen de Weerd, Rineke Verbrugge (2019). Training the use of theory of mind using artificial agents. *Journal on Multimodal User Interfaces*, *13*(1), 3-18.

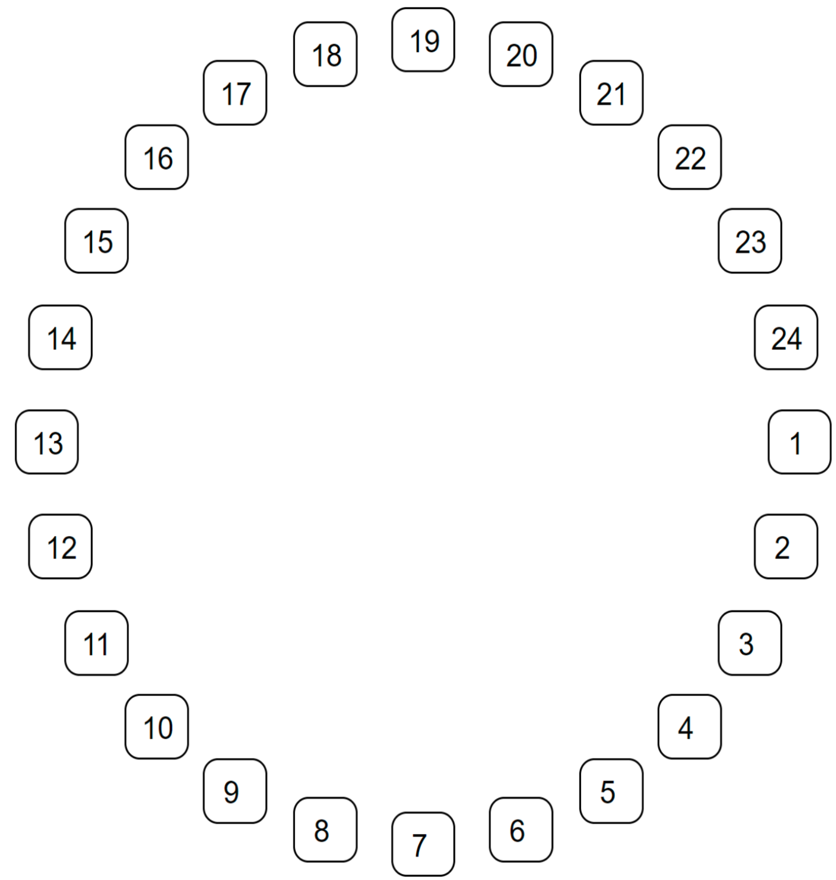
Methodology for investigating theory of mind in games

- Agent-based computational models
  - Simulate interacting agents
  - Introduce differences in the ability to use theory of mind
  - Compare performance among agents: Do higher orders of theory of mind allow agents to achieve better outcomes?
  - *Previous result*: For several competitive one-shot games such as rock-paper-scissors as well as some turn-taking games, 2nd order ToM appears optimal

-Verbrugge, R. (2009). Logic and social cognition: The facts matter, and so do computational models. *JPL* 38, 649–680.
-de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, *199*, 67-92.

The Mod game: rules

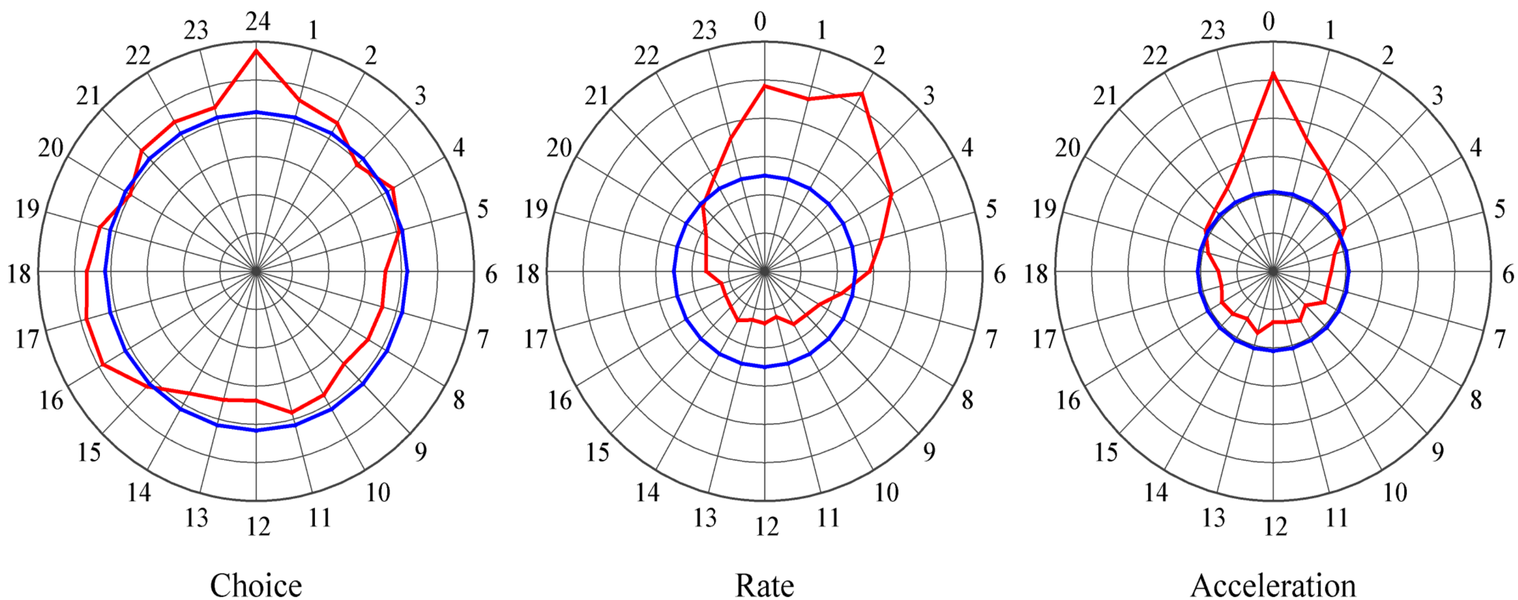
- 24 numbers arranged in a circle
- 2 player simultaneously choose a number
- You win a point if you choose the number that is exactly 1 higher than the number chosen by your opponent
  – Number 1 wins if the opponent chose 24



Frey, S. & Goldstone, R. L. (2013). Cyclic game dynamics driven by iterated reasoning. *PloS one*, *8*(2), e56416.

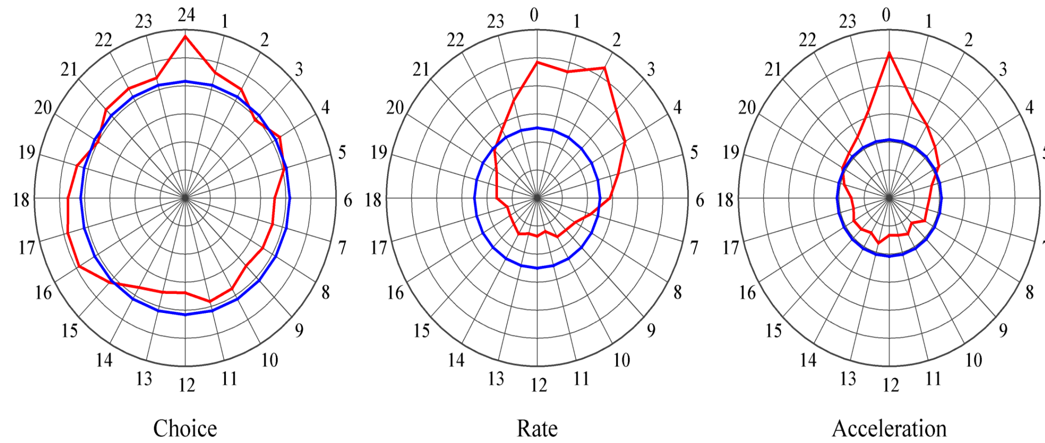
The Mod game: Nash equilibrium

- The only Nash equilibrium is random play, even in repeated play
- In repeated play, humans deviate from the Nash equilibrium
  - Participants tend to choose the number 2 higher than their previous choice



Frey, S. (2013). *Complex collective dynamics in human higher-level reasoning; A study over multiple methods* (PhD thesis, Indiana University).

The Mod game

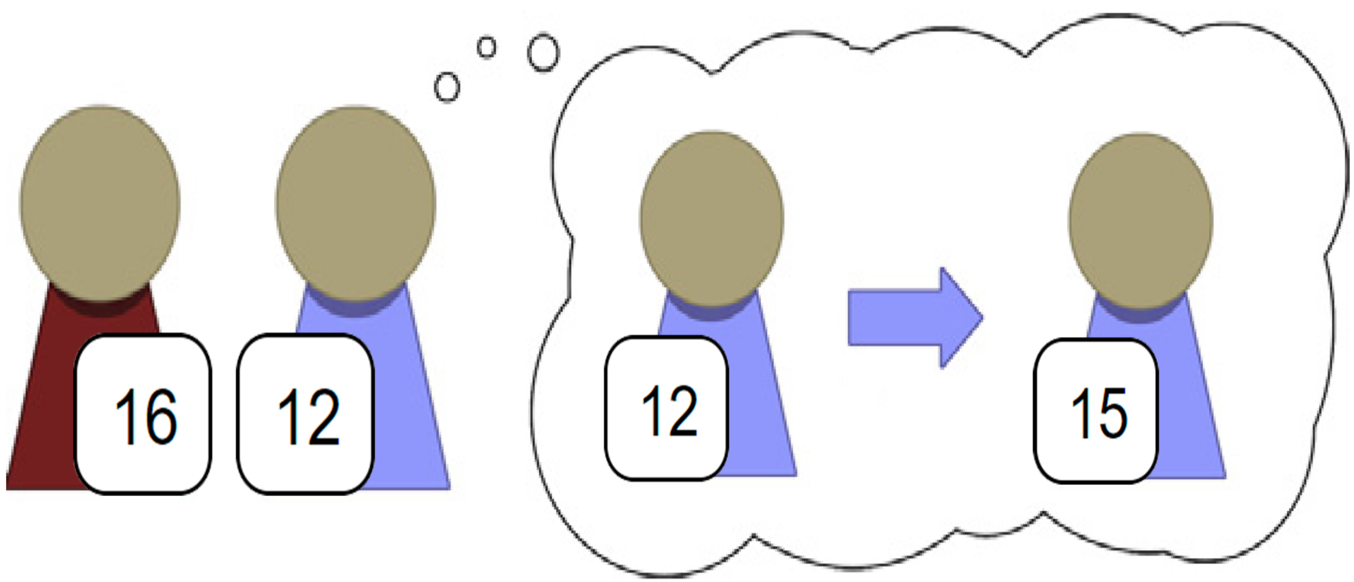


- What strategies do people use when playing the Mod game?
  - Are they only taking their most recent action into account?
  - Do they only react on the most recent action of the opponent?
  - Are they using theory of mind?

- Can people be encouraged to use higher-order theory of mind while playing the Mod game?

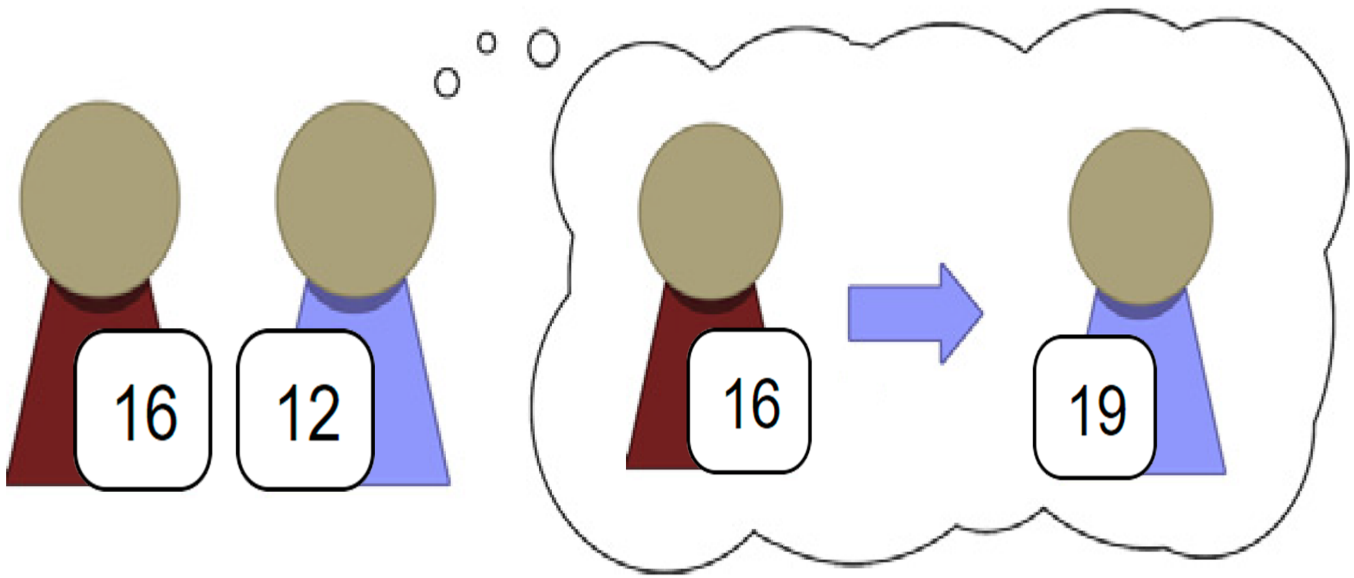
Possible strategies

- *k*-self-regarding strategy:
  - Almost always plays *k* higher than their own last action, for some *k*
  - Example of a 3-self-regarding strategy:



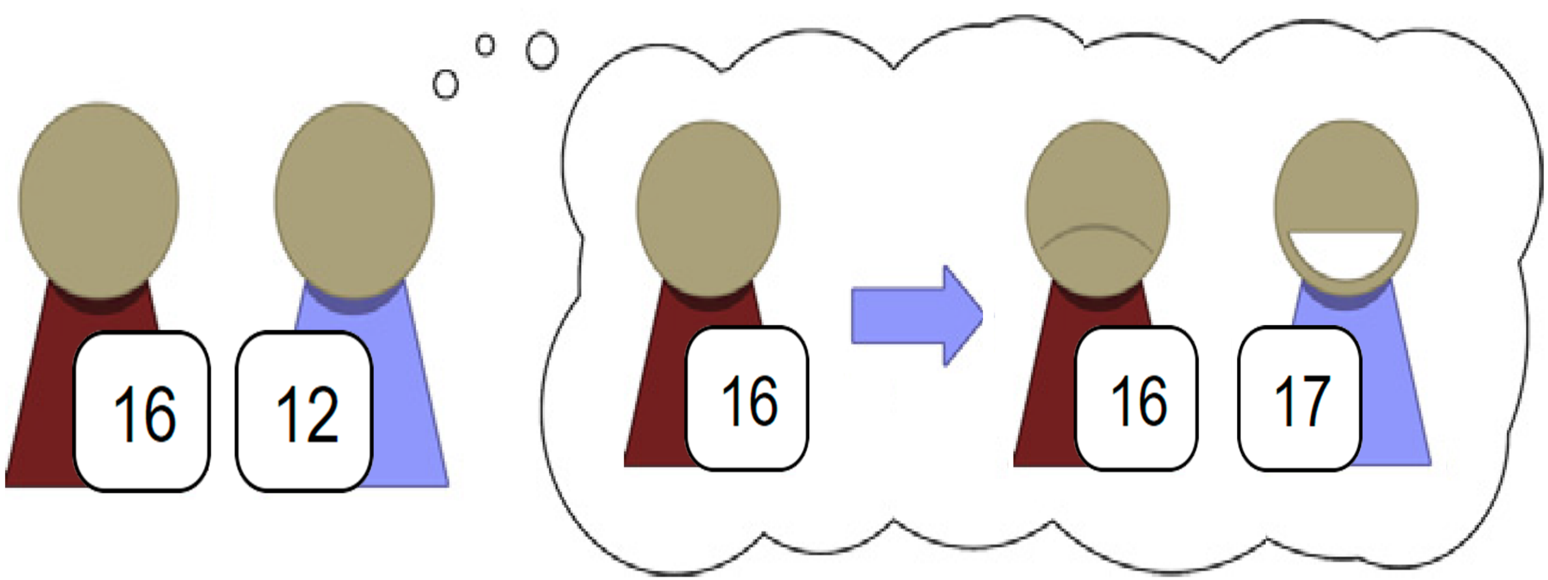
Possible strategies

- *k*-other-regarding strategy:
  - Almost always plays *k* higher than the opponent's last action, for some *k*
  - Example of a 3-other-regarding strategy:



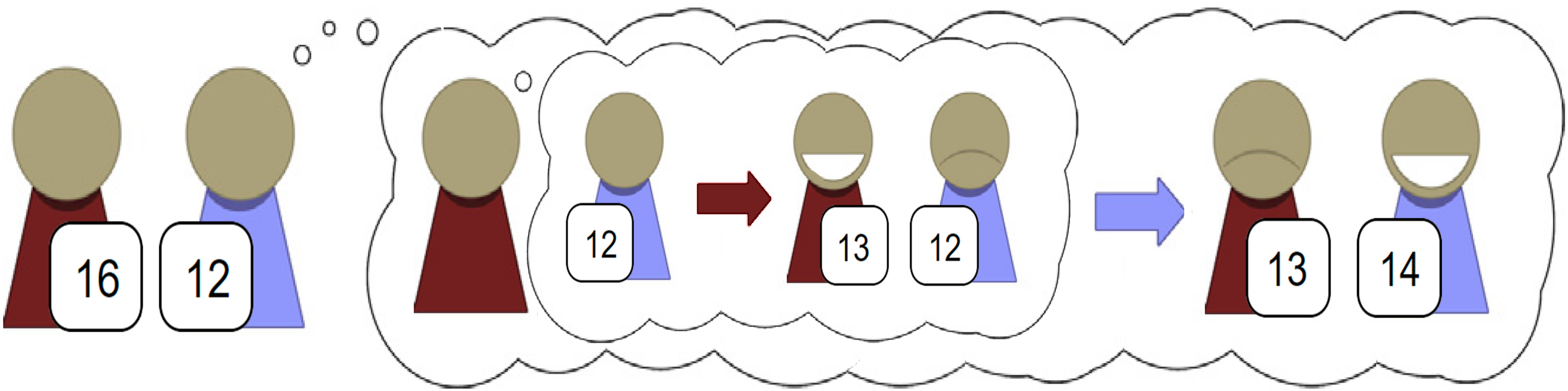
Possible strategies

- Zero-order theory of mind (ToM_0):
  - An opponent that chooses number *n* is likely to play that number again in the future



Possible strategies

- First-order theory of mind (ToM_1):
  - The opponent may also believe that I am a ToM_0 agent
  - A ToM_1 agent considers two predictions of opponent behavior



Strategies

- How can we distinguish between ToM_1 and self-regarding 2?
  - Both tend to choose 2 higher than their own previous choice
  - Self-regarding 2 deviates from this choice randomly
    - Independent of the choice of the opponent
    - Randomly distributed across the other 23 options
  - ToM_1 deviates from this choice predictably
    - In response to some unexpected behavior of the opponent
    - Non-uniformly distributed across the other 23 options

Experimental setup

- Participants play 8 blocks of 20 rounds each of the Mod game
- Participants play 2 blocks each against:
  - A ToM_1 agent
  - A ToM_2 agent
  - A ToM_3 agent
  - An agent that switches randomly between the other agents at each turn
- Participants are informed which agent they are playing against

Bayesian Model Selection

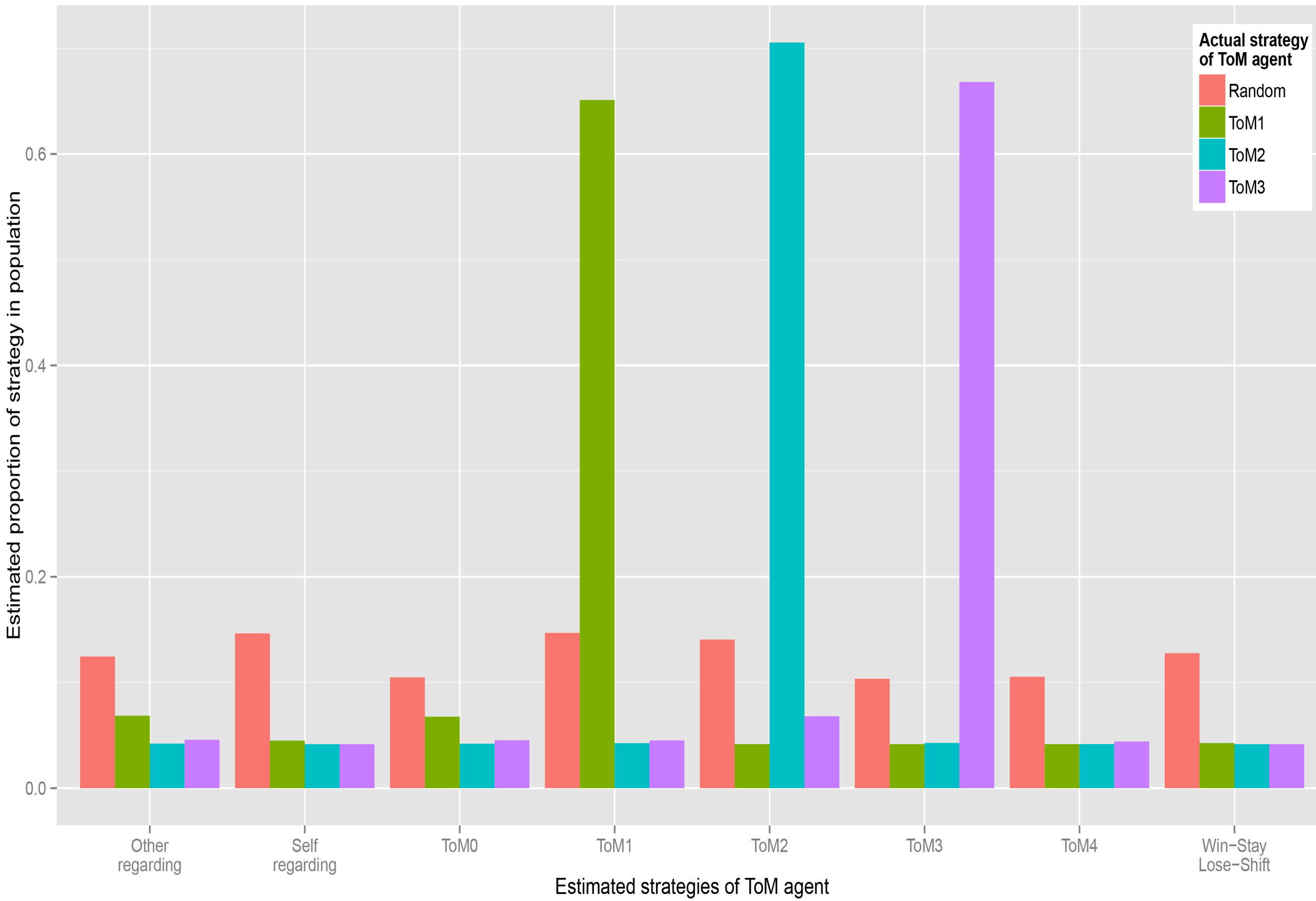
- Each strategy is a model of participant behavior
  - For each action *a*, the strategy *s* specifies the probability *P(a|s)* that a participant following that strategy *s* would perform action *a*
  - Using Bayes' rule, you can determine the likelihood *P(s|a)* of the participant using strategy *s*, given his or her actions *a*
- Assume **all** participants use **the same strategy**
  - Select the strategy that has the highest likelihood across all participants

Random-effects Bayesian Model Selection

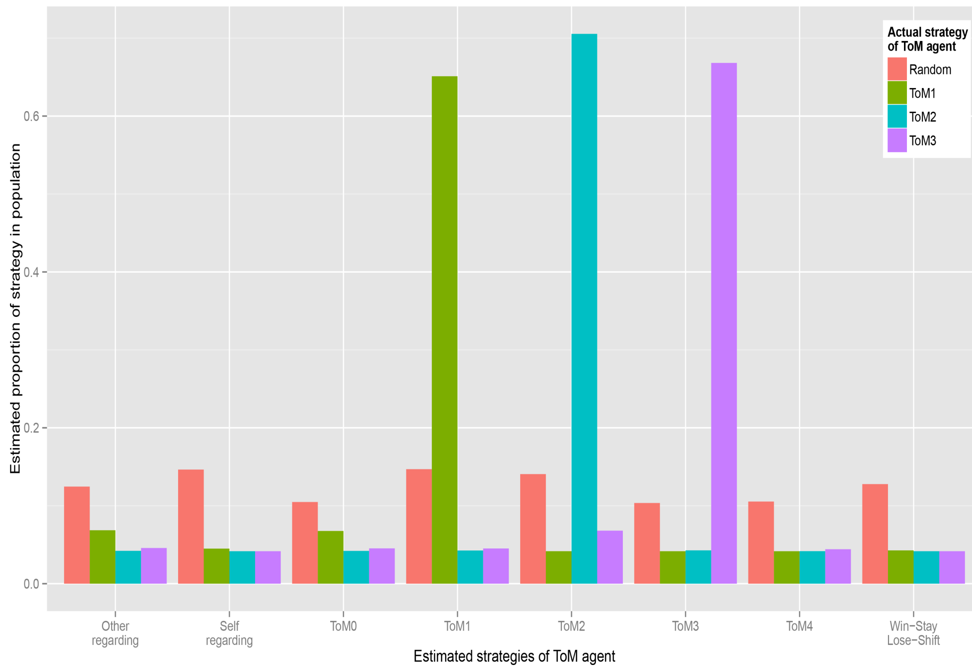
- Each strategy is a model of participant behavior
  - For each action *a*, the strategy *s* specifies the probability *P(a|s)* that a participant following that strategy *s* would perform action *a*
  - Using Bayes' rule, you can determine the likelihood *P(s|a)* of the participant using strategy *s*, given his or her actions *a*
- Assume participants **are selected randomly from a population of strategies**
  - Select the **distribution** of strategies that has the highest likelihood for the sample

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J. and Friston, K.J., 2009. Bayesian model selection for group studies. *Neuroimage*, *46*(4), pp.1004-1017.
de Weerd, H., Diepgrond, D. and Verbrugge, R., 2017. Estimating the use of higher-order theory of mind using computational agents. *The BE Journal of Theoretical Economics*, *18*(2).

Random-effects Bayesian Model Selection – Agent strategies

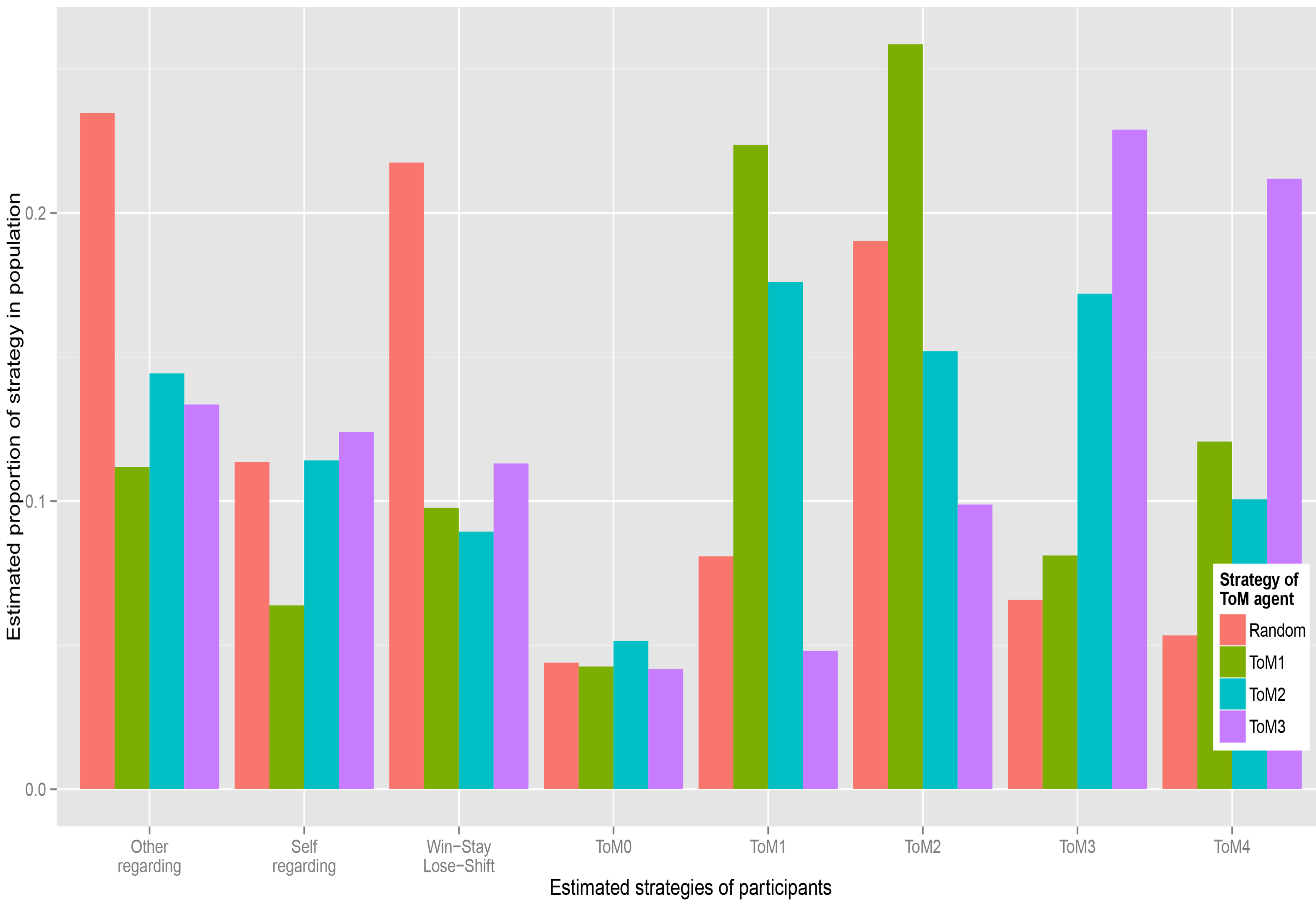


Random-effects Bayesian Model Selection – Agent strategies

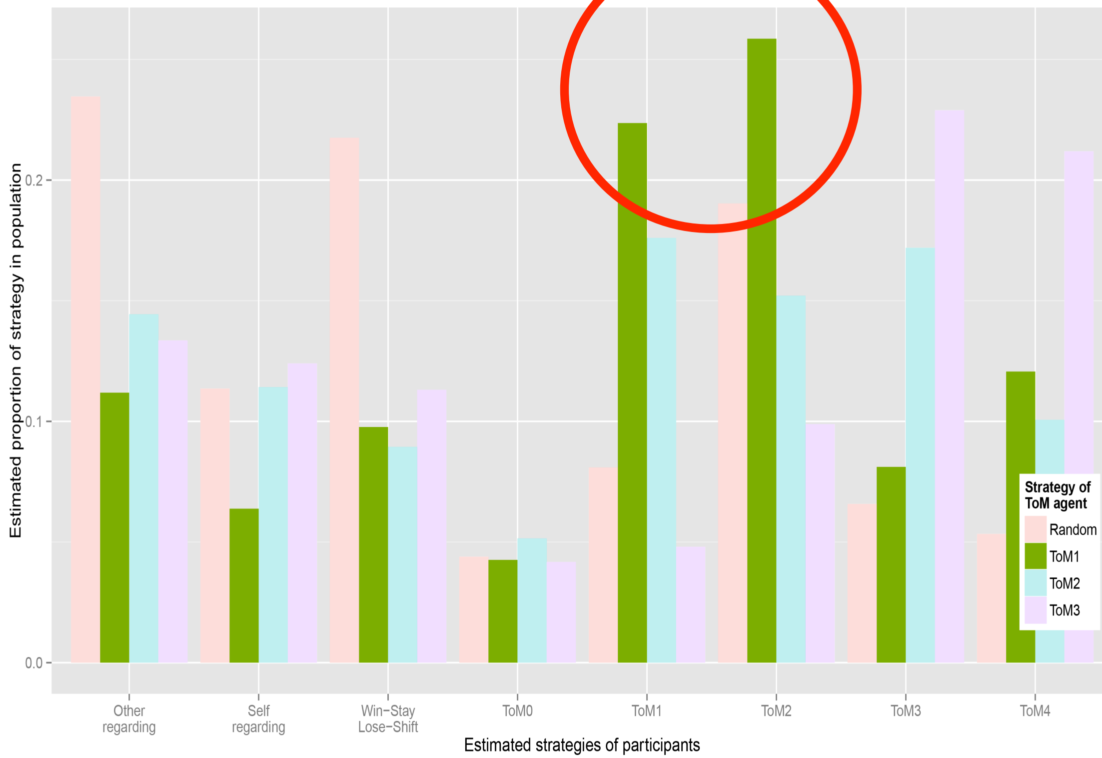


- ToM agents are identified as such
  - That is, agents are not underestimated or overestimated
- The randomizing agent is not recognizable as any strategy

Random-effects Bayesian Model Selection – Participant strategies

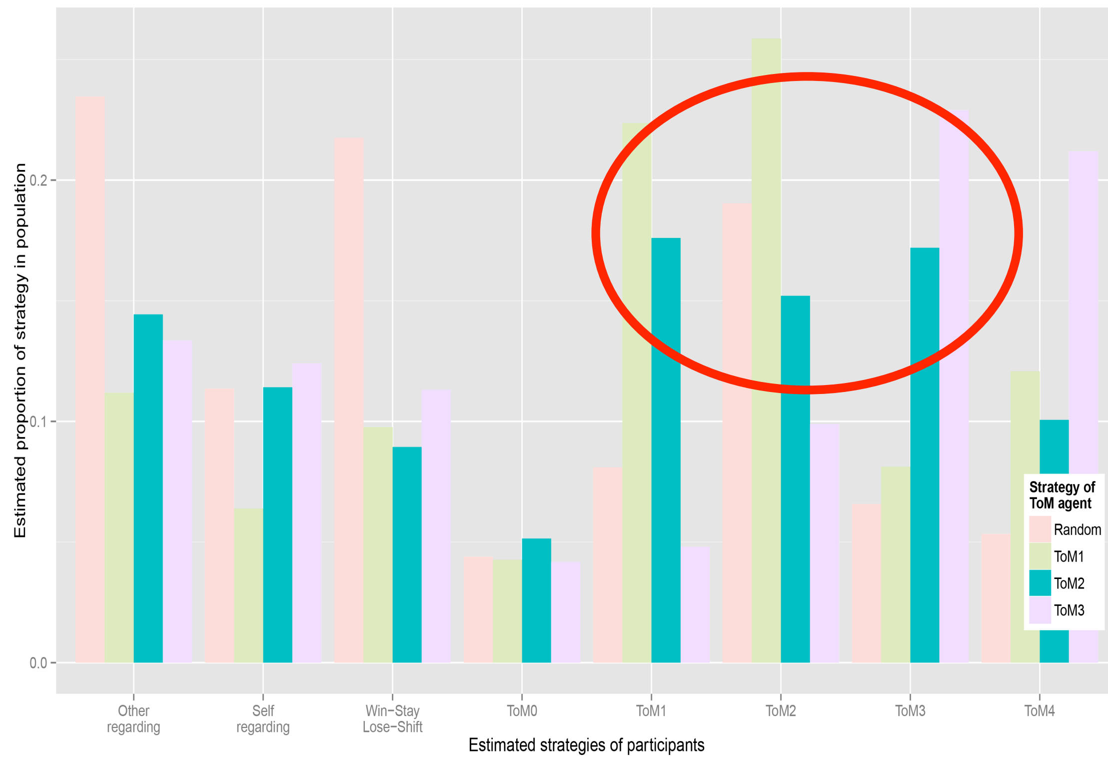


Random-effects Bayesian Model Selection – Participant strategies



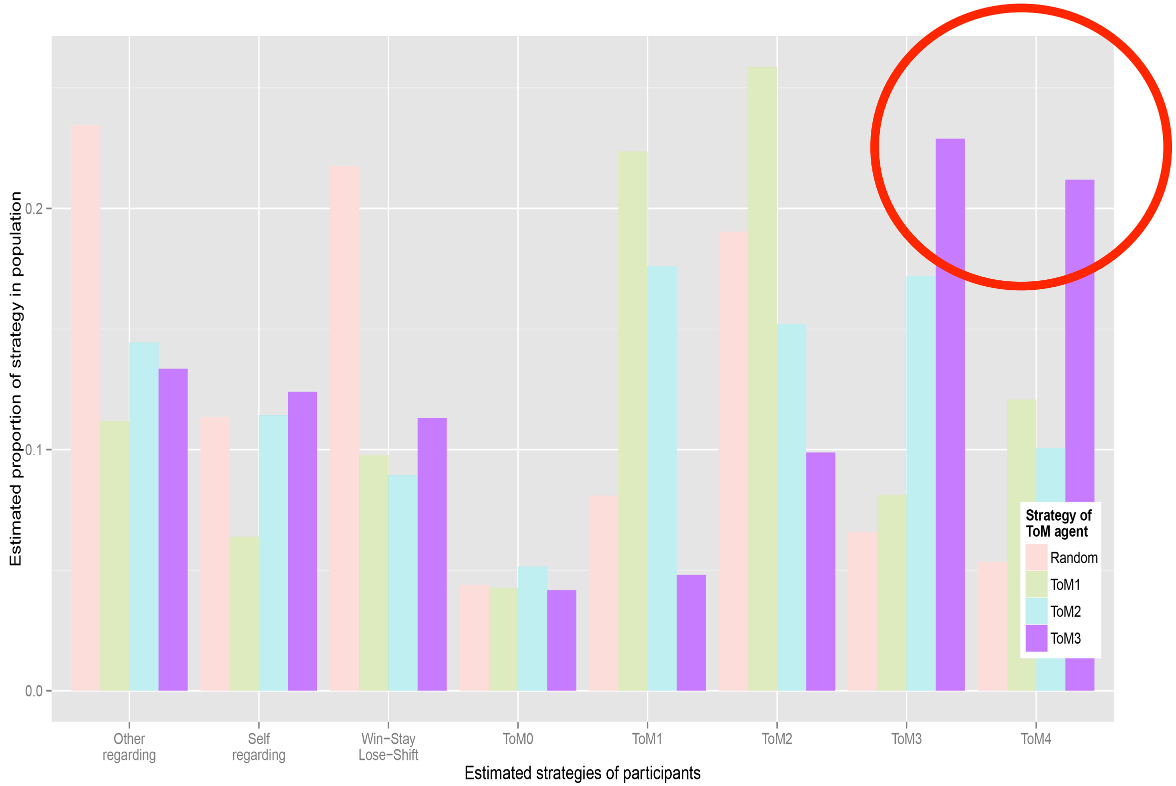
- When playing against a ToM_1 agent, participants mainly use first-order or second-order theory of mind

Random-effects Bayesian Model Selection – Participant strategies



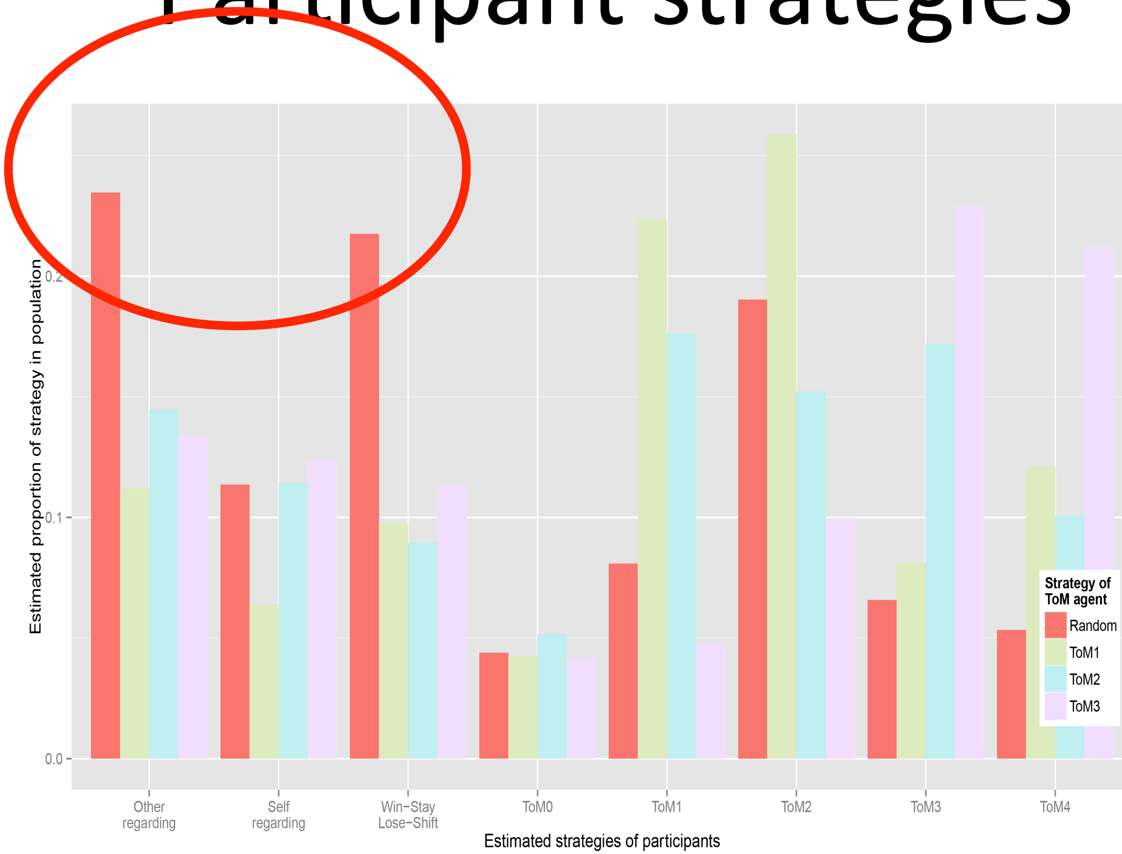
- When playing against a ToM_2 agent, participants use a variety of theory of mind strategies

Random-effects Bayesian Model Selection – Participant strategies



- When playing against a ToM_3 agent, participants mainly use third-order or fourth-order theory of mind

Random-effects Bayesian Model Selection – Participant strategies



- When playing against the randomizing agent, participants mainly make use of non-theory of mind strategies

Conclusions on the Mod game

- Participant behavior is best described as theory of mind strategizing, not as following simpler behavior-based strategy
  – unless the opponent behaves unpredictably

- Participants adapt their behavior in response to opponents
  – When playing against higher-order ToM agents, participants also use higher orders of ToM reasoning

- Random Effects Bayesian Model selection appears to be a good method to diagnose players’ strategies in iterated single-shot games where new choices depend on the history of choices

III. Negotiating with software agents

- Negotiations are situations with *mixed motives*, where participants have cooperative goals (to make a deal) & competitive goals (to get the most out of a trade)

- Agent-agent simulation result: second-order theory of mind is beneficial for agents in a negotiation game

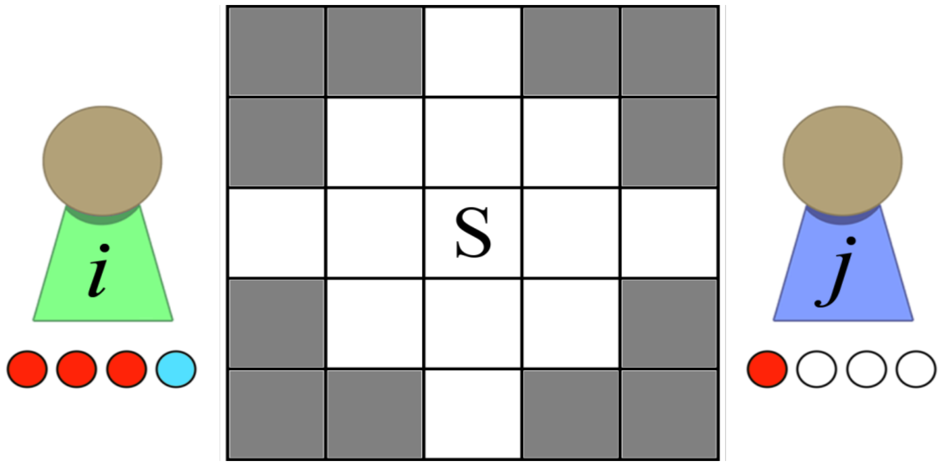
- Do students spontaneously use theory of mind in negotiation?

Methodology for investigating theory of mind in a negotiation game

- Behavioral experiments
  - Let participants play against theory of mind agents
  - Use a higher-order theory of mind agent to determine to what extent human participants use ToM and whether they dynamically adapt their level to their opponent's use of ToM

-Weerd, H. de, Broers, E., & Verbrugge, R. (2015) Savvy software agents can encourage the use of second-order theory of mind by negotiators. *Proceedings 37th Annual Meeting of the Cognitive Science Society*, pp. 542–547.

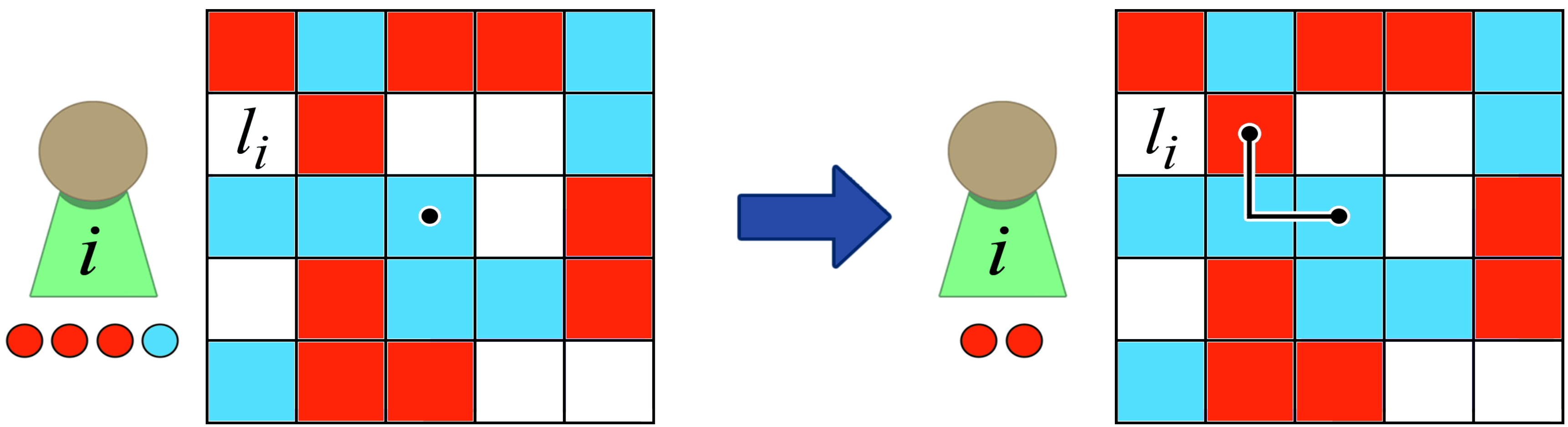
Colored trails: Negotiation game outline



- Each player has an initial location, goal location and set of chips
  - Each agent starts at the central square (marked 'S')
  - Goal locations are assigned randomly (gray squares)
  - Agents know their own goal location, but do not know the goal location of their trading partner (imperfect information)

Grosz, B., Kraus, S., Talman, S., Stossel, B., Havlin, M.
The influence of social dependencies on decision-making: Initial investigations with a new game. *Proceedings AAMAS* 2004

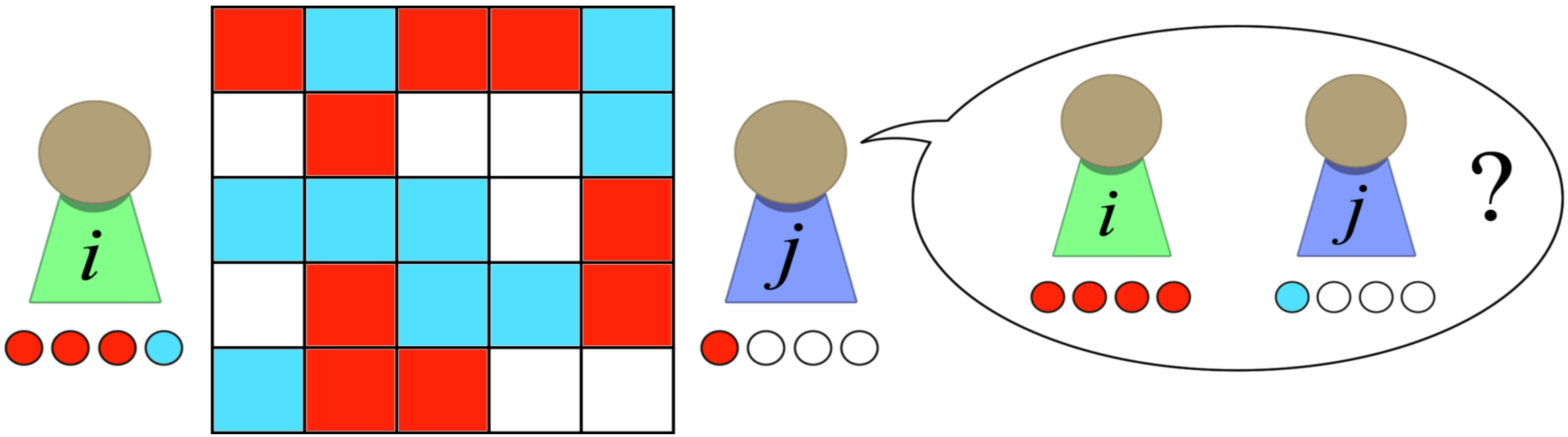
Colored trails: Scoring



- A player can move to an adjacent tile by handing in a chip of the same color as the destination tile
- Players are scored based on their final location:
  - Each step towards the goal is worth 100 points
  - Reaching the goal is worth an additional 500 points
  - Unused chips are worth 50 points each

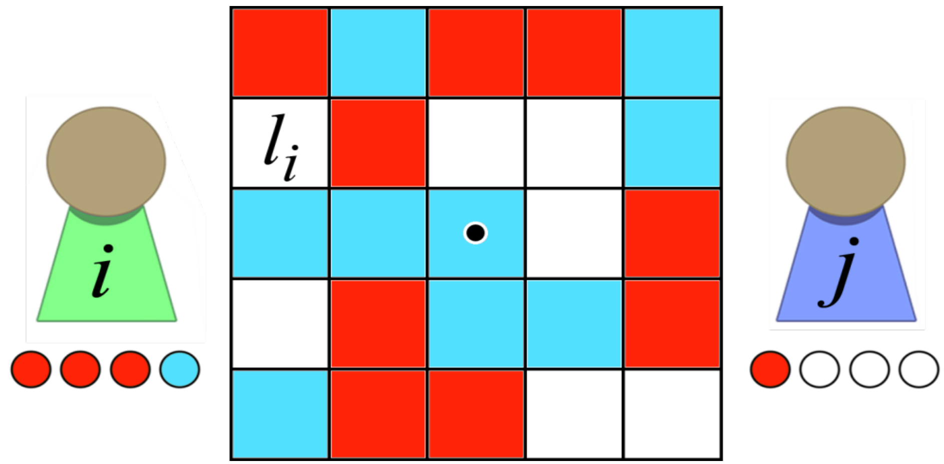
Here, initial score of *i* is 300 pts: 2 steps closer to goal + 2 chips left

Colored trails: Negotiation



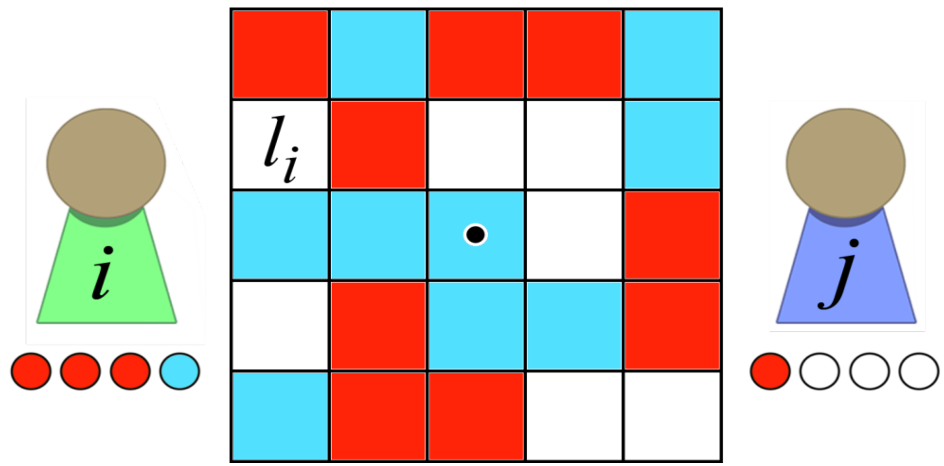
- Players alternate in offering a redistribution of chips:
  - Negotiation continues as long as agents make offers; repeating an offer is allowed
  - For each round of negotiation, agents pay a cost of 1 point
  - Negotiation succeeds if an offer is accepted
  - Negotiation fails if a player withdraws from negotiation; then each player's chips remains as originally allocated to them

Zero-order theory of mind agent *i*



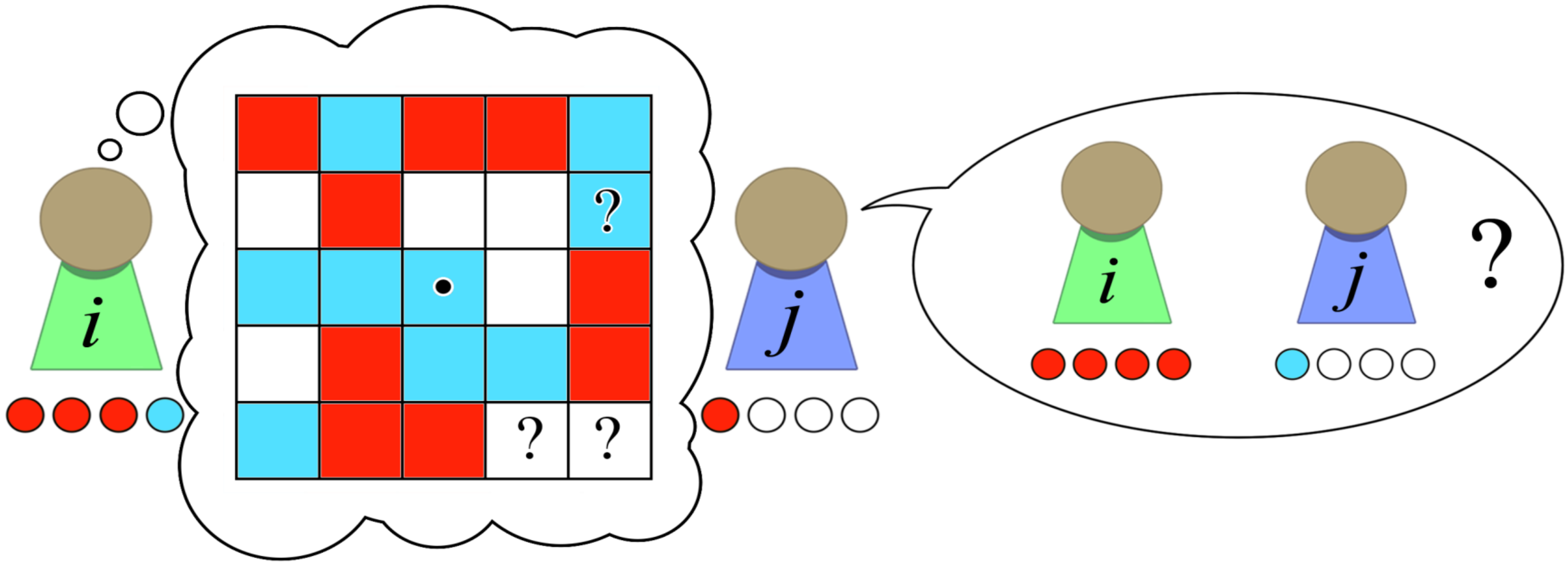
- If agent *i* is a zero-order theory of mind (ToM_0) agent, then *i* does not consider agent *j*'s goals
  - Instead, ToM_0 agent *i* only considers overt behavior, i.e., offers
  - Agent *i* forms beliefs about which offers agent *j* will accept, which are based on successful and failed offers in the past

First-order theory of mind agent *i*



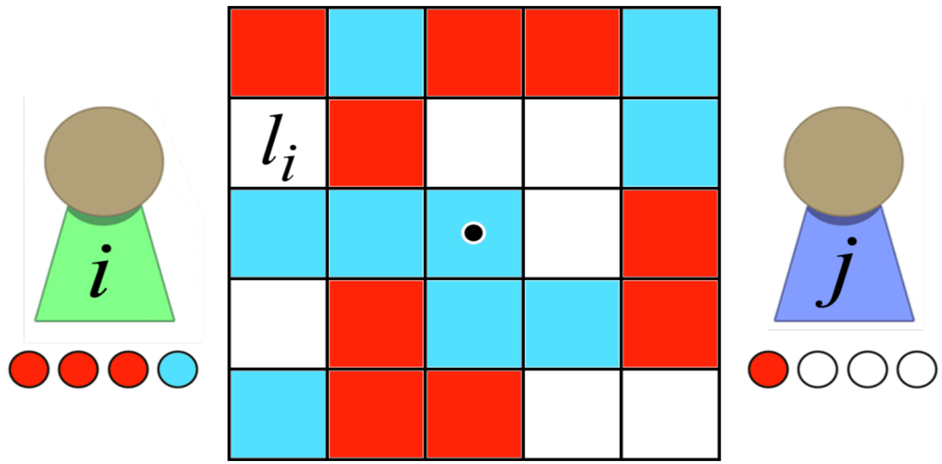
- A first-order theory of mind (*ToM_1*) agent reasons explicitly about his trading partner's goals
  - A *ToM_1* agent puts himself in the position of his partner and determines by simulation what he would have done in his partner's place

First-order theory of mind agent *i*, example



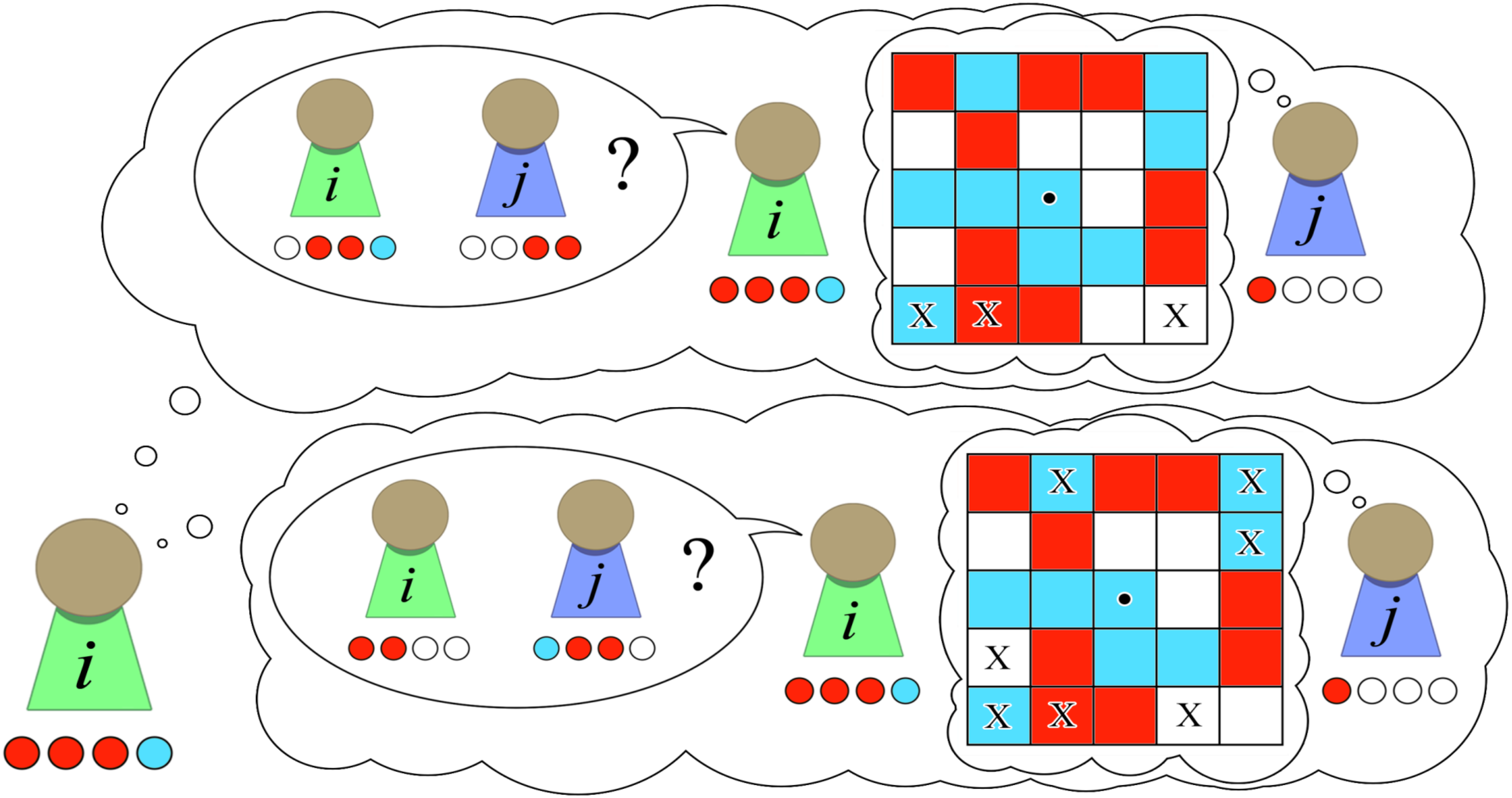
- Whenever agent *j* makes an offer, agent *i* learns about the goal of agent *j*
  - "Since agent *j* asks for the blue chip and offers his red one in return, agent *j* probably needs a blue chip, but not a red one, to reach his goal"

Second-order theory of mind agent *i*



- A ToM_2 agent *i* reasons about what agent *j* believes about agent *i*'s goals and beliefs:
  - Agent *i* believes that agent *j* tries to find out the goal location of agent *i*
  - Agent *i* can construct his offer in such a way to inform agent *j* about his own goal location

Second-order ToM agent *i*, example



Agent *i* can determine how much information he gives agent *j* about his goal location

Results of second-order ToM agents in simulations

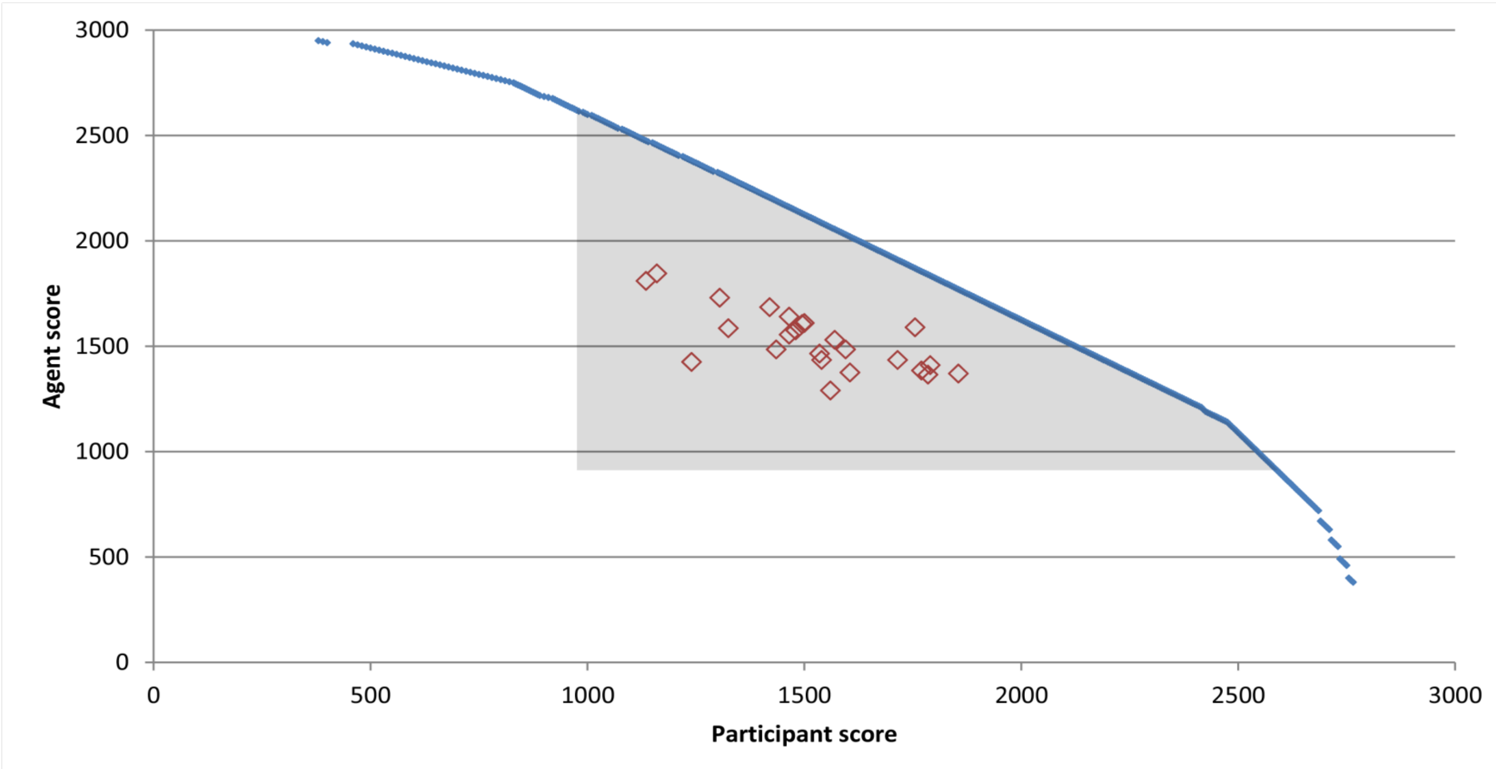
- ToM_2 agents outperform ToM_1 agents:
  - When a ToM_2 agent and a ToM_1 agent negotiate, the ToM_2 agent obtains at least as large a piece of the pie as their trading partner
- Two ToM_2 agents work well together:
  - When two ToM_2 agents negotiate, they typically split the pie into two equal pieces
  - Individual and collective incentives align, so behavior that yields a ToM_2 agent his highest gain also leads to highest collective performance

de Weerd, H., Verbrugge, R., & Verheij, B. Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information. *J. Autonomous Agents and Multi-Agent Systems*, 31(2): 250–287, 2017.

Experiment on negotiations of students with ToM_0, ToM_1, & ToM_2-agents

- Human participants play 24 Colored Trails games against computational agents
- Games are split up into 3 blocks of 8 games each
    - In each block, the theory of mind ability (ToM_0, ToM_1, ToM_2) of the computer player is different
    - Participants are *not* told about ToM agents, nor that the computer agent changes
    - Participants have one minute to decide on each action (offer, accept, or withdraw)
    - A negotiation game usually takes 4-6 rounds of offers and counteroffers

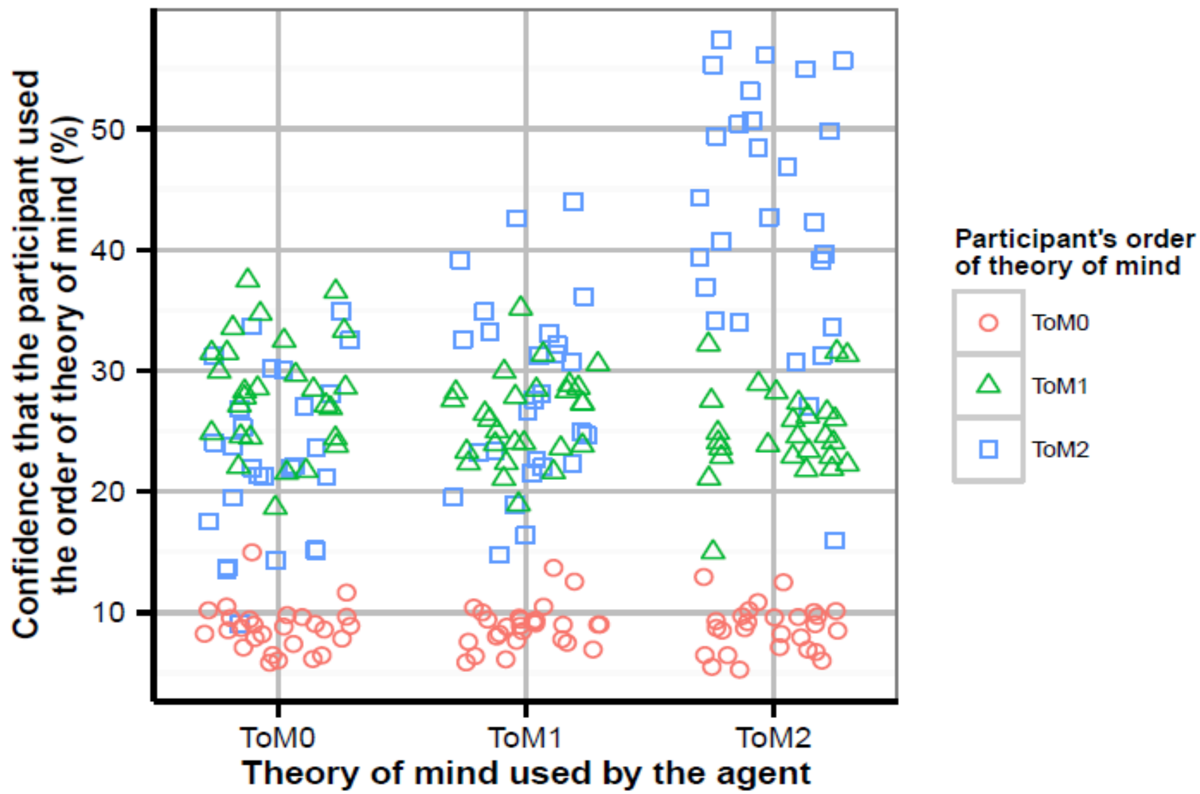
Participant performance over all 24 games



Human subjects and agents usually come to win-win agreements; their scores do not differ significantly

Weerd, H. de, Broers, E., & Verbrugge, R. (2015) Savvy software agents can encourage the use of second-order theory of mind by negotiators. *Proc 37th Annual Meeting of the Cognitive Science Society*, pp. 542–547.

Orders of theory of mind used by participants as estimated by a *ToM*$_3$ agent



Participants are classified as a mix of *ToM*$_1$ and *ToM*$_2$. Against *ToM*$_2$ agents, participants act more like *ToM*$_2$ agents

Student participants & theory of mind agents in the negotiation game

- Participants spontaneously use ToM_1, ToM_2 when they negotiate with agents
- A ToM_3 software agent can estimate, based on a number of different negotiation games, whether the participant plays ToM_0, ToM_1, or ToM_2
    - but it cannot discern other strategies.
- Participants adjust their ToM level to their partner
    - They made more ToM_2 offers when paired with ToM_2 agents

Conclusions

- It is possible to entice people to use higher levels of theory of mind in several types of game:
  - (im)perfect information,
  - one-shot or turn-taking,
  - competitive or mixed-motive negotiation
- For different types of game, apply different methods to estimate a human participant's reasoning strategy

Future research

- Maintaining lies, detecting lies, require 2nd-order ToM
  – Build computational cognitive models of non-cooperative communication, simulating participants in experiments
- Develop serious games to train people in complex social skills such as negotiation
  – Build training partners for children, adults, & people with autism spectrum disorder
- And finally: create Hybrid Intelligence
  – Enable mixed teams of software agents, robots and human beings, where strengths of all can be combined,
  – Members model one another's mental states and agents can estimate human members' ToM abilities.