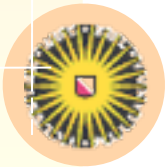


Logica voor Artificiële Intelligentie

Part II, Lecture 8

Wrapping it up

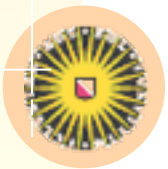


Universiteit Utrecht

J.M.Broersen@uu.nl

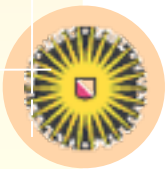
Subjects of today

- The 7 cards
- 1000 gems



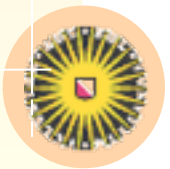
Problem description

- three agents: A, B, and C
- seven cards, numbered 1...7.
- A and B have three cards, C has one card
- the agents know how many cards (from the set of seven) other agents have, but they do not know which ones.
- how can agents A and B communicate publicly, in such a way that the two of them will commonly know (1) the exact distribution of cards, (2) that C does not get to know for any card except its own in whose possession it is?
- Interesting aspect: an example about secrecy, i.e., about how to **avoid** common knowledge.



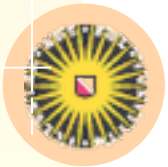
A concrete instance of the problem

- Agent A has cards $\langle 1, 2, 3 \rangle$
- Agent B has cards $\langle 4, 5, 6 \rangle$
- Agent C has card $\langle 7 \rangle$



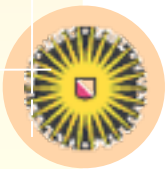
Possible worlds structure

- Agents A and B each consider 4 possible worlds.
- Agent C considers 20 possible worlds.
- Agents A and B have to take advantage of the fact that they have more knowledge.
- Public communications can only lead to an **increase in knowledge**, and thus, **less possible worlds**.



How to view knowledge exchange?

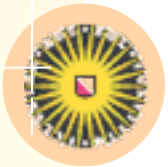
- If, in a public communication, an agent acquires new knowledge, the result is always the elimination of some of the worlds it considered possible.
→ any knowledge exchange is **equivalent** with the exclusion of some set of possible worlds



How to talk about possible worlds?

- Example: ‘the sum of my cards is 6’ is the same as ‘my cards are $\langle 1,2,3 \rangle$ ’
- Example: ‘the sum of my cards is 8’ is the same as ‘my cards are $\langle 1,2,5 \rangle \vee \langle 1,3,4 \rangle$ ’

Conclusion: *Any message* can only convey information equivalent to some announcement of the form ‘my cards are $\langle x_1, y_1, z_1 \rangle \vee \langle x_2, y_2, z_2 \rangle \vee \langle x_3, y_3, z_3 \rangle \dots$ ’



Simplifying assumption

- We only look for solutions where first A, in one public announcement, conveys information that results in complete knowledge for B.
- Then, for the second public communication, it suffices that B announces what card C has (by which C learns nothing, and A learns everything).



Criteria for A's announcement

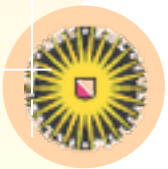
Without loss of generality, we assume that A's announcement is of the form:

'my cards are $\langle 1,2,3 \rangle \vee \langle x_1, y_1, z_1 \rangle \vee \langle x_2, y_2, z_2 \rangle \dots$ '

The $\langle x_i, y_i, z_i \rangle$ should be such that:

1. A and B commonly get to know the distribution.
2. C does not learn the position of **any** card but its own, a condition that is commonly known to A and B.

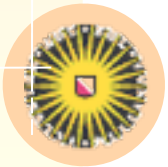
I will refer to the $\langle x_i, y_i, z_i \rangle$ as 'wrong' disjuncts.



Hints

So:

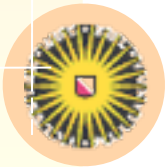
1. The disjunction should be sufficiently small, in order to reveal A's cards to agent B.
2. The disjunction should be sufficiently large, in order not to reveal the place of any of the cards to agent C.



Are there solutions?

There are 102 (!) different solutions:

- 60 solutions with 5 disjuncts
- 36 solutions with 6 disjuncts
- 6 solutions with 7 disjuncts



One solution for the given situation

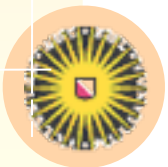
A announces: ‘my cards are $\langle 1,2,3 \rangle \vee \langle 1,5,7 \rangle \vee \langle 2,4,7 \rangle \vee \langle 2,5,6 \rangle \vee \langle 3,4,6 \rangle$ ’.

B’s possible worlds are: $\langle 1,2,3 \rangle \vee \langle 1,2,7 \rangle \vee \langle 1,3,7 \rangle \vee \langle 2,3,7 \rangle$

- B takes the intersection and concludes: $\langle 1,2,3 \rangle$!
- **A knows that B knows.** A does not know that C’s card is 7. But whatever C’s card is, ‘wrong’ disjuncts in A’s announcement contain **at least 2** ‘wrong’ cards. So A knows that B knows that these are wrong: among two wrong cards there is at least one possessed by B, because maximally one is owned by C.

A knows that C does not get to know the position of any card. A knows that C can rule out only 2 of the 4 wrong disjuncts, because any card A does not possess appears exactly 2 times in the wrong disjuncts. For the particular card distribution of the above example: the intersection with A’s announcement is: $\langle 1,2,3 \rangle \vee \langle 2,5,6 \rangle \vee \langle 3,4,6 \rangle$, which means that C does not know the position of any of the cards but its own!

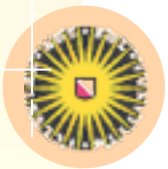
- Question: does C get to know anything?
→ yes, because he goes from 20 to only 3 possible worlds.



Student solution nr.1

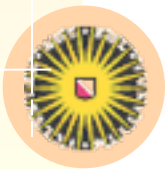
Poging 1: A stuurt 8 mogelijkheden: de 4 die hij zelf voor mogelijk houdt plus 4 die 'swaps' zijn daarvan (de set kaarten van A en B verwisseld). C ziet dan nog 2 opties, B maar 1!

Poging 2: Een oplossing met een randomgenerator.



Student solution nr.2

A zegt: "Ik heb hand $\langle a,b,c \rangle$ of $\langle a,d,e \rangle$ of $\langle b,e,f \rangle$ of $\langle c,f,g \rangle$."



Student solution nr.3

Persoon A zegt tegen persoon B:

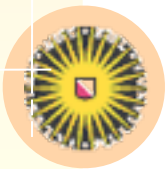
Ik heb:

(ofwel (a&b) ofwel (d&e) ofwel (e&f) ofwel (f&g))

EN ik heb:

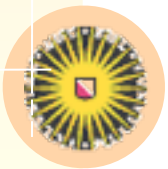
(ofwel (b & c) ofwel (c&e) ofwel (d&f) ofwel (e&g))

n.b. 'ofwel' is dus een xof.



Student solution nr.4

"Ik heb ABC, ADE of BFG."



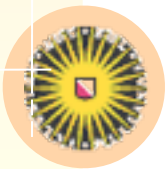
Wrong claim nr. 1

- A's wrong disjuncts **must** contain at least (and thus exactly) one of his own cards!?

First, if that is true our solution would not work (!):

1. C knows that $\langle 1, 5, 7 \rangle$ is wrong. So C would conclude that A has either 1 or 5 (and that B has the other one of this pair).
2. So, in the example, C could rule out $\langle 3, 4, 6 \rangle$.

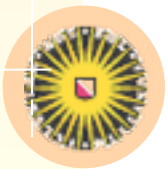
But then C would know that $\langle 1, 2, 3 \rangle \vee \langle 2, 5, 6 \rangle$ and thus that A has 2!



So what is an example?

$\langle 1,2,3 \rangle \vee \langle 4,5,6 \rangle \vee \langle 1,4,7 \rangle \vee \langle 2,5,7 \rangle \vee \langle 3,6,7 \rangle$.

- Only 5 disjuncts.
- The wrong disjunct $\langle 4,5,6 \rangle$ contains none of the correct cards
- C's card appears 3 times!
- So C sees only 2 possible worlds (A has $\langle 1,2,3 \rangle$ and B $\langle 4,5,6 \rangle$, or the other way around), but still does not know the position of any card!
- Does not depend on this particular distribution of the cards and thus not on C actually having card 7!



Wrong claim nr. 2

“One solution is simply to take for the wrong disjuncts all 4 possible permutations of the other cards” (was eerder een tentamenvraag)

$\langle 1,2,3 \rangle \vee \langle 4,5,6 \rangle \vee \langle 4,5,7 \rangle \vee \langle 4,6,7 \rangle \vee \langle 5,6,7 \rangle$.

At first, all seems ok. A and B learn the cards, C will consider (in the actual situation) $\langle 1,2,3 \rangle \vee \langle 4,5,6 \rangle$ possible, so does not know a card...

- So why is this still a wrong solution?!?
- Well, by clever reasoning C learns all cards! He sees that B cannot have $\langle 1,2,3 \rangle$ because in that case B would not be able to deduce that A would have $\langle 4,5,6 \rangle$.
- A has given **too many** wrong disjuncts containing none of his cards...



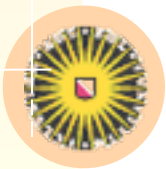
Universiteit Utrecht

J.M.Broersen@uu.nl

Comment

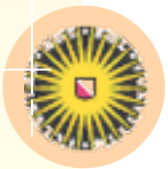
So why is it that A cannot just add more and more wrong disjuncts (= hands, = possible worlds) to his announcement? After all as long as the disjuncts contain maximally one of his own cards, B will always be able to exclude them. And C will only get more distracted by them. Or not?

No, since C knows that all disjuncts except will have to be excluded by B and thus need to contain at least one of B's cards, the more wrong disjuncts A adds, the more C gets to know about B's cards.



What about doing this in PAL/BMS?

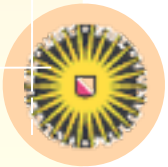
- PAL + ‘group common knowledge’ is enough, this is about public announcements and conditions of common knowledge for sub-groups.
- However, we can use PAL (extended with group common knowledge) to **check** our solutions, not to **find** one (except for the non-attractive procedure of checking all possible solutions).



Example 2: the 1000 gems

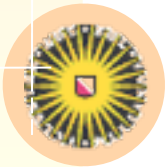
- There are n agents, and 1000 gems.
- Agent n gets an assignment: ‘make a proposal to all other $n-1$ agents for a division of the 1000 gems. If more than 50% of the agents (including yourself) agrees, the division is executed. Otherwise you are killed, and the division assignment is given to agent $n-1$ ’.
- Agents want to stay alive, get as much gems as possible, and kill if that does not cost them gems.
- Agents take no risks.
- Agents are perfectly rational and know this of each other.

If $n=5$, what will agent 5 propose, and why (how does he reason)?



Initial consideration

- The first impression is that an agent can never propose a division that will be accepted, because other agents want him to be killed, such that more gems are left for them. In particular agent 1 wants to kill, since he never will be killed himself...
- So, is there a solution at all?



The solution from backwards induction

Proposal of x to y	1	2	3	4	5
2	Any x , because 1 kills 2 anyhow	$1000-x$, (but killed)			
3	0	0	1000		
4	1	1	0	998	
5	2 0	0 2	1 1	0 0	997 997

Do things change if we add that agents do **not** kill for pleasure?
 (= aspect opponent model!)

Universiteit Utrecht

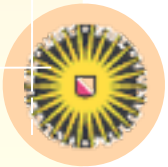
J.M.Broersen@uu.nl



Questions

What happens if we add another agent? How does agent 6 deal with the uncertainty at level 5?

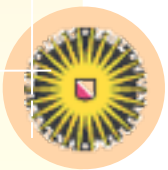
If an agent at level 5 can either be offered 0 or 2, will it accept 2 at level 6 just to avoid the risk of being offered nothing at level 5?



The solution from backwards induction when pleasure kills are absent

Proposal of x to y	1	2	3	4	5
2	1000	0			
3	0	1	999		
4	1	2	0	997	
5	2	0	1	0	997

Recall: agents take no risks... Therefore, agent 3 proposes 1 instead of 0 to agent 2.

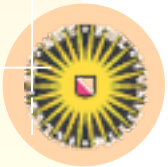


Universiteit Utrecht

J.M.Broersen@uu.nl

Questions / discussion

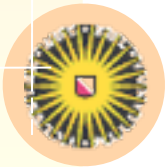
- What if an agent does **not accept** an offer that is calculated according to backwards induction?
- Might it be that this is a **'threat'** to the agents higher in the chain that he needs more gems than follows from the BI-reasoning?
- Could that be rational play?
- Would we not reason: the spoiler might want more, but we do not know how many, so better to give him nothing and count on satisfying others we expect to obey BI.
- Does this give a paradox?



Universiteit Utrecht

J.M.Broersen@uu.nl

Epistemic interaction



Universiteit Utrecht

J.M.Broersen@uu.nl

Epistemic reasoning with wrong information

- *A traveling salesman found himself spending the night at home with his wife when one of his trips was accidentally cancelled. The two of them were sound asleep, when in the middle of the night there was a loud knock at the front door. The wife woke up with a start and cried out, ‘Oh my God! It’s my husband!’ Whereupon the husband leapt out of bed, ran across the room and jumped out the window. [Schank and Abelson, 1977, p. 59.]*
- Conclusion: man and wife both cheat + both have wrong beliefs about who they are in bed with

