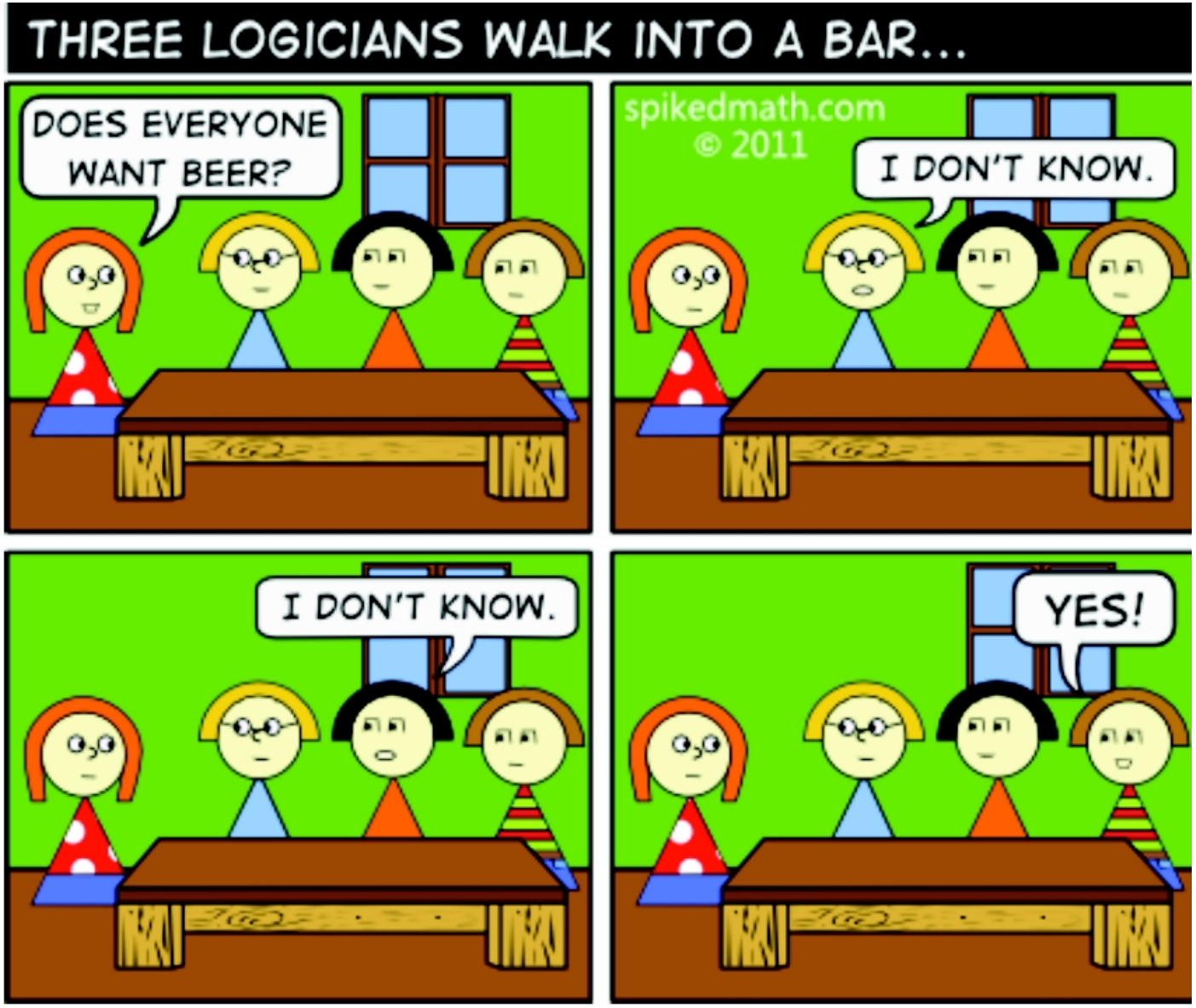
Reasoning About Knowledge

Adrian Groza

Department of Computer Science
Technical University of Cluj-Napoca

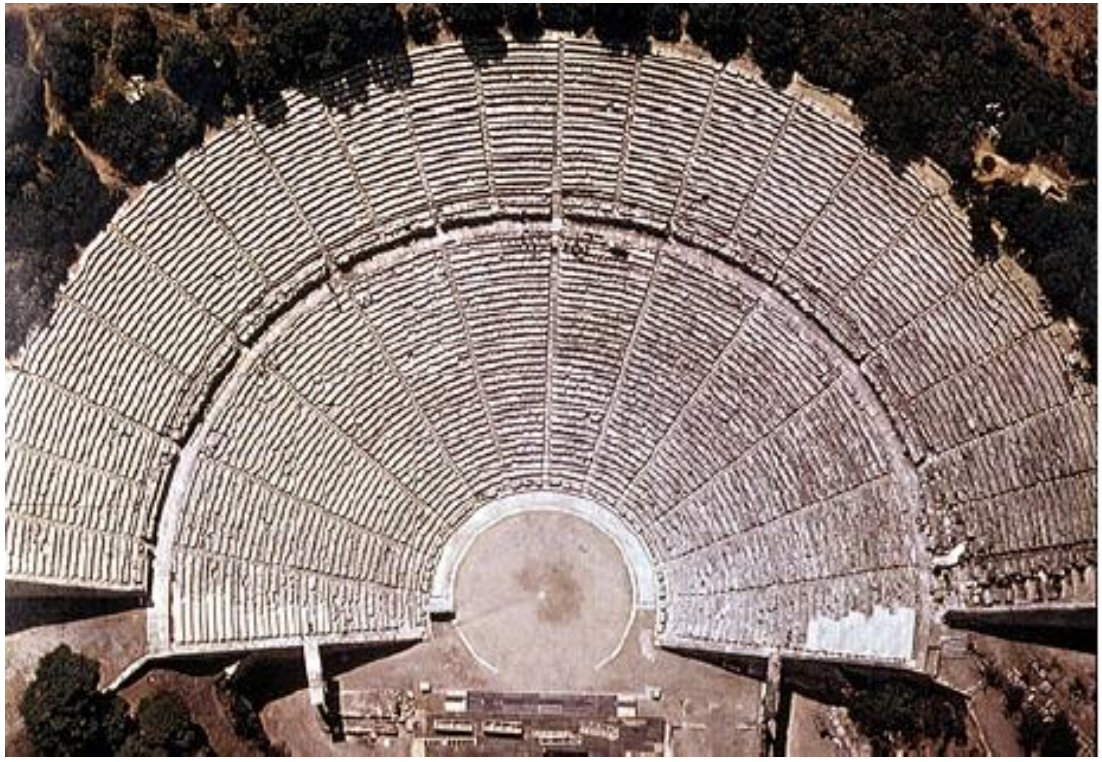
Outline

Warm-up


THREE LOGICIANS WALK INTO A BAR...
DOES EVERYONE WANT BEER?
spikedmath.com
© 2011
I DON'T KNOW.
I DON'T KNOW.
YES!

Persons in the play:

- the logician, the king
- the computer scientist
- the cognitive scientist
- the philosopher
- the economist
- Abelard
- Heloise
- 4 muddy children
- 2 sport bidders
- a byzantine general
- a soldier



Common Experience: The theatre of Epidaurus

- Time: Spring, 7 weeks before graduation.
- Scene: room B519 at TUKE, eager students

Epistemic Logic

Example (What does Anne know?)

Anne draws one from a stack of three different cards 0, 1, 2.
She draws card 0. She does not look at her card yet!
Card 1 is put back into the stack holder.
Card 2 is put (face down) on the table.
Anne now looks at her card.

- Anne holds card 0.
- Anne knows that she holds card 0.
- Anne does not know that card 1 is on the table.
- Anne considers it possible that card 1 is on the table.
- Anne knows that card 1 or card 2 is in the stack holder.
- Anne knows her own card.

Descriptions of Knowledge

- There is one agent Anne: $\{a\}$
- Propositional variables q_a for 'card q is held by Anne.'
- $K_a\varphi$ expresses 'Anne knows that φ'
- $\hat{K}_a\varphi$ expresses 'Anne considers it is possible that φ' $(\neg K_a \neg \varphi)$

- Anne holds card 0: 0_a
- Anne knows that she holds card 0: $K_a 0_a$
- Anne does not know that card 1 is on the table: $\neg K_a 1_t$
- Anne considers it possible that card 1 is not on the table: $\neg K_a \neg 1_t$ (she does not know that not 1_t)
- Anne knows that card 1 or card 2 is in the stack holder: $K_a(1_h \vee 2_h)$
- Anne knows her own card: $K_a 0_a \vee K_a 1_a \vee K_a 2_a$

What is epistemic logic about?

Episteme (Greek) = knowledge

I know that p	$K_a p$
He does not know that p	$\neg K_b p$
He knows whether p	$K_b p \vee K_b \neg p$
He knows that I know that she does not know that p	$K_b K_a \neg K_c p$

Language

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi$$

Alice knows that Bob knows that Alice knows p, and Bob does not know that Alice knows that Bob does not know that Alice knows p

The Properties of Knowledge (Axioms)

1. Distribution axiom (K): if an agent knows φ and knows that $\varphi \rightarrow \psi$, then the agent must also know ψ.
   $K_a\varphi \wedge K_a(\varphi \rightarrow \psi) \rightarrow K_a\psi$
2. Truth axiom (T): if an agent knows facts, the facts must be true (knowledge implies veracity). $K_a\varphi \rightarrow \varphi$
3. Knowledge generalization rule (N): if φ is true in every world that an agent considers to be a possible world, then the agent must know φ at every possible world $M \models \varphi$ then $M \models K_i\varphi$
4. Positive introspection (KK): agents know what they know $K_a\varphi \rightarrow K_aK_a\varphi$
5. Negative introspection: agents know what they do not know $\neg K_a\varphi \rightarrow K_a\neg K_a\varphi$

Logical Omniscience

$K_a\varphi \wedge K_a(\varphi \to \psi) \to K_a\psi$

Our epistemic agent *a* knows all the logical consequences. If *Q* is a logical consequence of *P*, then there is no possible world where *P* is true but *Q* is not.

Example

If *a* knows that prime numbers are divisible only by themselves and the number one then *a* knows that 86833176188118864955181944012799999999 is prime.

Possible World Semantics

- Divide the set of possible worlds between those that are compatible with an agent's knowledge, and those that are not.
- $K_a\varphi$: in all possible worlds compatible with what a knows, it is the case that φ
- To express the idea that for agent a, the world w' is compatible with his information state, or accessible from the possible world w which a is currently in, it is required an accesability relation $R(\rightarrow)$ to hold between w and w'.
- The arrow $w \rightarrow w'$ means that, if one is living in the alternate world w, then w' is one of the imaginary worlds that he would think of as possible.
- Reflexivity says that there is an arrow $w \rightarrow w'$ from every world to itself. Reflexivity means that the actual world is always one of the worlds that is imagined as possible.

Kripke Model

Definition

A *Kripke model* is a structure $M = \langle S, R, V \rangle$, where

1. *domain* S is a nonempty set of states (or possible words);
2. R yields an *accessibility relation* $R_a \subseteq S \times S$ for every $a \in A$ (it is meant to capture what worlds or states agent i considers to be possible);
3. *valuation* (function) $V : P \to 2^{|S|}$ an interpretation function that determines which sentences in the languages are true in which worlds (states)

- If all the relations R_a in M are *equivalence relations*, we call M an epistemic model. In that case, we write $\sim_a$ rather than R_a, and we represent the model as $M = \langle S, \sim, V \rangle$.
- The truth assignment tells us whether or not a proposition $p \in P$ is true or false in a certain state. Just because something is true in one world does not mean it is true in another. To show that a formula φ is true at a certain world, one writes $(M, s) \models \varphi$, normally read as "φ is true at (M, s)" or "(M, s) satisfies φ"

Epistemic modeling

- Given is a description of a situation, the modeler tries to determine:
  - The set of relevant propositions P
  - The set of relevant agents A
  - The set of states S
  - An indistinguishability relation over these states for each agent R_a

Example

- $S = \{012, 021, 102, 120, 201, 210\}$
- $\sim_a = \{(012, 012), (012, 021), (021, 021), \dots\}$
- $V(0_a) = \{012, 021\}$, $V(1_a) = \{102, 120\}$, ...

$Hexa1, 012 \models K_a 0_a$

$\Leftrightarrow$

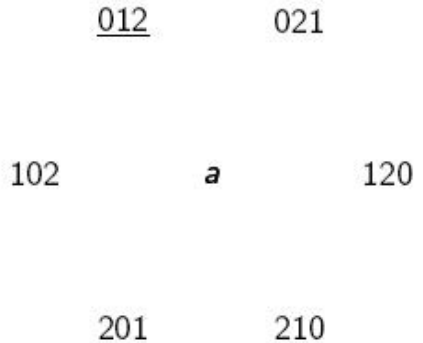
for all $t : 012 \sim_a t$ implies $Hexa1, t \models 0_a$

$\Leftarrow$

$Hexa1, 012 \models 0_a$ and $Hexa1, 021 \models 0_a$

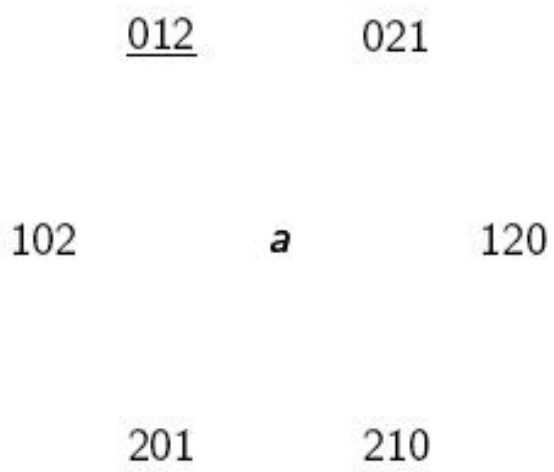
$\Leftrightarrow$

$012 \in V(0_a) = \{012, 021\}$ and $021 \in V(0_a) = \{012, 021\}$



Two agents

Anne and Bill draw 0 and 1 from the cards 0, 1, 2.
Card 2 is put (face down) on the table.



- Bill does not consider it possible that Anne has card 1: $\neg \hat{K}_b 1_a$
- Anne considers it possible that Bill considers it possible that she has card 1: $\hat{K}_a \hat{K}_b 1_a$
- Anne knows Bill to consider it possible that she has card 0: $K_a \hat{K}_b 0_a$

The Wise Persons Puzzle

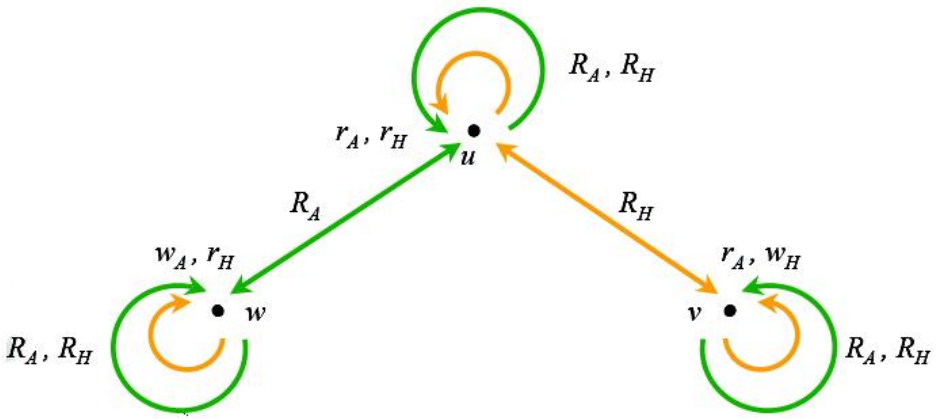
Participants: Abelard (A), Heloise (H), the King
It is common knowledge among them that:

- There are three hats: 2 red hats and 1 white hat
- The King places a hat on each of A's and H's heads
- A and H cannot see their own hat, but
- A and H can see the other person's hat

The following discussion now takes place:
King: "Abelard, do you know the color of your hat?"
Abelard: "No"
King: "Heloise, do you know the color of your hat?"
H: "Yes"
Question: What is the color of Heloise's hat?

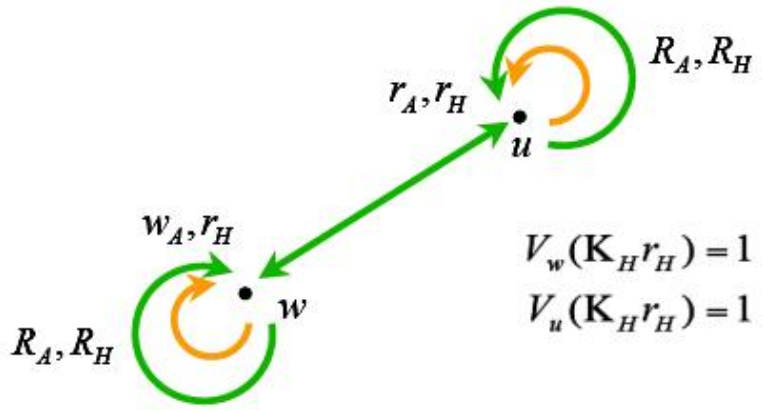
Epistemic Analysis

r_A: Abelard wears a red hat; r_H : Heloise wears a red hat
w_A : Abelard wears a white hat; w_H: Heloise wears a white hat



$K_A r_A$ is true in v but false in u
$K_A w_A$ is false in w

King: ”Abelard, do you know the color of your hat?”
A: No



Less accessibility arrows corresponds to less ignorance, thus more knowledge

Truth

- An atomic propositional formula, p, is said to be true in a world s iff s is in the set of possible worlds assigned to p

$$M, s \models p \text{ iff } s \in V(p)$$

- The formula $K_a\varphi$ is true in a world s

$M, s \models K_a\varphi$ iff for all t such that $s \sim_a t$ it holds that $M, t \models \varphi$

From Knowledge to Belief

- Veridity axioms: $K_a\varphi \to \varphi$: it is impossible to know something that is not true.
- Logic of belief: $B_a\varphi$: agent a believes φ is true

Example

If George knows that Alice believes it is raining, then Alice knows that Bob knows it.

$K_G B_A rain \to K_A K_B B_A rain$.

- Knowledge implies belief: $K_i\varphi \to B_i\varphi$.

Outline

Conventions and Common Knowledge

Philosopher: Imagine driving on a one-lane road. You just came out of the Channel Tunnel on the British side. Drivers who just went from France to England tend to forget on which side of the road they have to drive, particularly if they find themselves on a quiet one-lane road where they are suddenly confronted with oncoming traffic. Then you will have to swerve a bit to the side to let it pass. In fact, you each have to swerve a bit. Will you swerve left or right?

Economist: Ah, this is beginning to sound familiar! If you swerve, you're a chicken. If not, and if you force the other to swerve, you're a tough guy. Unfortunately, when two tough guys come together, they will crash. There is interesting equilibrium behavior in examples like this. It's a standard setting for a two-person game in game theory.

Philosopher: Yes, you are right, but that is not what I wanted to explain. (To the cognitive scientist again:) Will you swerve left or right?

Cognitive Scientist: If I remember that I am in England, I will swerve left. Otherwise, I will swerve right.

Philosopher: Yes, and how about the guy coming towards you? He and you may both be cautious drivers, but if he swerves right and you left, you will still crash.

Conventions and Common Knowledge

Philosopher: It is not enough for you and the on-comer both to know that you have to drive left. You would also like to know that the other knows. And this will affect your behavior. You are very cautious if you do not know, slightly less cautious if you know but not if the other knows, even less cautious if you know and also know that the other knows but not if he knows that, and so on: you become more and more confident about the other's road behavior but never entirely so. Driving on the left-hand side is a convention, and this means that everyone knows that everyone knows know that everyone knows... up to any finite stack of knowledge operators.
Logician: Exactly, that's common knowledge.

Achieving Common Knowledge

Computer Scientist: Common knowledge is often easily achieved, by means of public announcement.
Cognitive Scientist: And what do you mean by public announcement, exactly?
Computer Scientist: A public announcement is an event where something is being said aloud, while everybody is aware of who is present, and it is already common knowledge that all present are awake and aware, and that everybody hears the announcement, and that everybody is aware of the fact that everybody hears it, and ...
Economist: Are e-mail notifications proper public announcements?

Message Exchange Cannot Create Common Knowledge

Logician: It was proved that message exchange in a distributed environment, where there is no guarantee that messages get delivered, cannot create common knowledge.

Computer Scientist: Analysis of message passing through unreliable channels is old hat in computer science. We call it the coordinated attack problem.

Two generals are planning a coordinated attack on a city. They are on two hills on opposite sides of the city, each with their own army, and they know they can only succeed in capturing the city if their two armies attack at the same time. But the valley that separates the two hills is in enemy hands, and any messengers that are sent from one army base to the other run a severe risk to get captured. The generals have agreed on a joint attack, but they still have to settle the time.

The Byzantine Generals

Picture by Marco Swaen

Open Secret

Computer Scientist: Indeed, here are those touchy situations where some proposition is common knowledge, but the participants mutually pretend that the contrary proposition is the case

> *You are celebrating St. Nicholas with family friends. How will you behave if its generally known that your 8-year old niece does not believe in St. Nocholas? What if she knows that you know? And if it is common knowledge?*

> *I know that I have dropped that potato and so do you; but I hope and I believe that you do not know, and you hope that I do not know that you know"*

Common Knowledge and the Public Arena

Cognitive Scientist: Yes, but how does one know that an announcement has become common knowledge? I might have let my attention wander for a moment.
Computer Scientist: If an announcement is made, you were supposed to pay attention, and therefore the information can now be assumed common knowledge.
Philosopher: That is what happens in the public arena all the time. At the basis of legal relations between individuals and the state, is the assumption that the law is common knowledge.
Cognitive Scientist: But this is a fiction. Professional lawyers have a full-time job to keep up with the law. Ordinary citizens can simply not be expected to cope.

Common Knowledge and the Public Arena

Philosopher: I prefer to say that it is a necessary presumption. Roman lawgivers found out long ago that if citizens within their jurisdiction could plead innocence because of being unaware of the law, no offender could ever get convicted. So they were quick to invent principles like *Ignorantia legis neminem excusat*, “ignorance of the law excuses no one”.

Computer Scientist: As a counterpart the laws have to be properly published and distributed. Of course, the citizens are not supposed to read all that boring stuff, but they should be able to find out about it whenever they want. In this way, the publications in the government gazette amount to public announcements.

General Knowledge

General knowledge $E\varphi$ *for set of agents* $\{1;...;n\}$:
$E\varphi := K_1\varphi \wedge K_2\varphi \wedge ... \wedge K_n\varphi$
Logician: General knowledge among the members of a group of agents means that all individuals in the group know a certain fact φ.
$E_G\varphi$ which reads every agent in group G knows φ;

Are $E\varphi$ and $EE\varphi$ equivalent?
More generaly
if $E^1\varphi = E\varphi$ and inductively $E^{k+1} = E(E^k)\varphi$ for $k \geq 1 \Rightarrow$
E^{k+1} and E^k are not in general equivalent

Logician: Common knowledge means: everybody knows φ everybody knows that everybody knows, and so on:
$C\varphi := \varphi \wedge E\varphi \wedge EE\varphi \wedge ...$
or
$C\varphi := \varphi \wedge K_1\varphi \wedge K_2\varphi \wedge K_1K_1\varphi \wedge K_1K_2\varphi \wedge \ldots K_1K_1K_1\varphi \ldots$
Computer Scientist: Let me propose a definition of common knowledge: A proposition φ is common knowledge if everybody knows that φ and everybody knows that φ is common knowledge.
$C\varphi \leftrightarrow \varphi \wedge EC\varphi$
Philosopher: That can hardly qualify as a definition, it's obviously circular.

Defining Common Knowledge by Co-recursion

Computer Scientist: "$C\varphi$ iff $\varphi \wedge EC\varphi$" is an instance of a definition by co-recursion. They are like recursive definitions, but with the crucial diference that there is no base case. And they define infinite objects. Let me give an example: An infinite stream of zeros, call it zeros, can be defined as: zeros equals a zero followed by zeros. In lazy functional programming this is written as
zeros = 0: zeros

Common Knowledge Seems Hard to Achieve

Philosopher: I am still wondering about this funny kind of definition that you call co-recursion. It seems like some kind of infinitary process is going on. How can we make sure it ever stops? I mean, imagine sending a romantic email, with “I adore you” or that sort of thing. You get a reply “I am so glad to know that you adore me”, you send a reply back “Now I am delighted, for I know that you know that I adore you”, only to get an exciting response: ”How sweet for me to know that you know that I know that you adore me.” Obviously, this nonsense could go on forever, and never achieve common knowledge of the basic romantic fact.

Logician: That’s brilliant. For it does never stop if you do it like this. But if the two lovebirds get together, they may still go through the whole exchange that you mentioned, but only for the fun of it. For the first “I adore you” creates common knowledge.

Co-presence Creates Common Knowledge

Philosopher: What are the properties of events that succeed in creating common knowledge? It seems to me that they all involve a shared awareness that a common experience takes place. It can involve various senses: hearing, eye-contact, maybe even touching or smelling. If B sees A look at B, then A sees B look at A. From this and a few simpler properties one can demonstrate that eye contact leads to common knowledge of the presence of the interactants. It is no coincidence that eye contact is of considerable emotional and normative significance

Co-presence Creates Common Knowledge

Example: cash withdrawal from a bank.

You withdraw a large amount of money from your bank account and have it paid out to you in cash by the cashier. The cashier looks at you earnestly to make sure she has your full attention, and then she slowly counts out the banknotes for you: one thousand (counting ten notes), two thousand (counting another ten notes), three thousand (ten notes again), and four thousand (another ten notes). This ritual creates common knowledge that forty banknotes of a hun- dred euros were paid out to you.

Philosophical question: when money is paid out to you by an ATM, does this create common knowledge between you and the machine?

Achieving Common Knowledge

Philosopher: Such rituals are important, indeed. Suppose you have four thousand bucks in an envelope, and you hand it over to a friend who is going to do a carpentry job at your home, say. Then what if this friend calls you later with dismay in his voice, and the message that there were just thirty-five banknotes in the envelope?

Economist: Then you are in trouble indeed, for you have failed to create common knowledge that the forty notes were there when you handed over the envelope. You failed to observe an important ritual, and this failure may result in the end of a friendship.

Knowledge in Groups

- Everybody knows individually
  - Example: Every family member knows that Saint Nicholas does not exist (but mother does not know that Miruna knows).
- Common knowledge
  - Everybody knows that p and
  - everybody knows that everybody knows that p
  - and ... etc.
  - Example: ”KBS class takes place on Thursday” is common knowledge among participants.
- Distributed knowledge
  - Members have different pieces of knowledge, e.g.
  - Example: Adriana knows lemma A. Vlad knows that lemma A implies theorem B. Adriana and Vlad have distributed knowledge of B.

Individual Ignorance vs. Common Ignorance

- Individual Knowledge about φ: $K_a\varphi \vee K_a\neg\varphi$
- Individual Ignorance about φ: $\neg K_a\varphi \wedge \neg K_a\neg\varphi$
- Common knowledge: $C\varphi \vee C\neg\varphi$
- Common ignorance: $\neg C\varphi \vee \neg C\neg\varphi$
- Commonly known common ignorance: $C(\neg C\varphi \vee \neg C\neg\varphi)$

Agreeing to Disagree

Economist: In the economics setting, instead of different possible situations - such as driving on the left, or on the right - the preferred model is that of different probable situations, and how events relate prior to posterior probabilities. In *"Agreeing to disagree"*, Aumann shows that if agents have common knowledge of their posterior probabilities of an event, that these must then be the same. In other words, they can only agree to agree and they cannot agree to disagree. It is not rational to agree to disagree, because this agreement would entail awareness of the fact that the disagreement can be exploited.

Agreeing to Disagree

Logician: What does it mean that you believe that the probability of an event is 1/2? Simply that if you are taking bets on this, then you will consider a bet with a return of two to one a fair bet. And if you believe that the probability is 1/4 and you are in a betting mood, then you will consider a bet with a return of four to one (including the stake) a fair bet.
Computer Scientist: That's what bookies call an odds of three to one against: If the event happens you win three times your stake, otherwise you lose your stake.

A Dutch Book about UEFA cup?

Will the winner be Manchester or Bayern Munchen?

- Adrian: probability the winner will be Bayern Munchen. Adrian is willing to take odds of one to one against Bayern Munchen.
- Alexandru: probability that Bayern Munchen will win is 1/4. Alexandru is willing to take odds of three to one against Bayern Munchen.

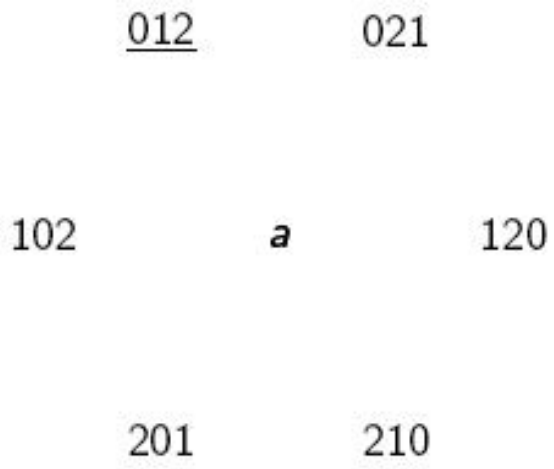
A Dutch Book about UEFA cup?

Computer Scientist: Assume a student places two bets:

- A bet of 1000- to Alexandru that Bayern Munchen will win (for three to one against)
- A bet of 2000 - to Adrian that Manchester will win (for one to one against).

If Bayern Munchen wins, the student collects 3000,- from Alexandru and loses 2000,- to Adrian: gain of 1,000.
If Manchester wins, the student loses her stake of 1,000 to Alexandru but collects 2,000 from Adrian: gain of 1,000.
A Dutch book is a set of odds and bets which guarantees a profit. Agreeing to disagree is not rational for people willing to take bets on their beliefs.

Example



- After Anne says that she does not have card 1, Cath knows that Bill has card 1.
  $[\neg 1_a]K_c 1_b$
- After Anne says that she does not have card 1, Cath knows Anne's card.
  $[\neg 1_a](K_c 0_a \vee K_c 1_a \vee K_c 2_a)$
- Bill still doesn't know Anne's card after that:
  $[\neg 1_a]\neg(K_b 0_a \vee K_b 1_a \vee K_b 2_a)$

Public Announcements

- Language

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid C_B\varphi \mid [\varphi]\varphi$$

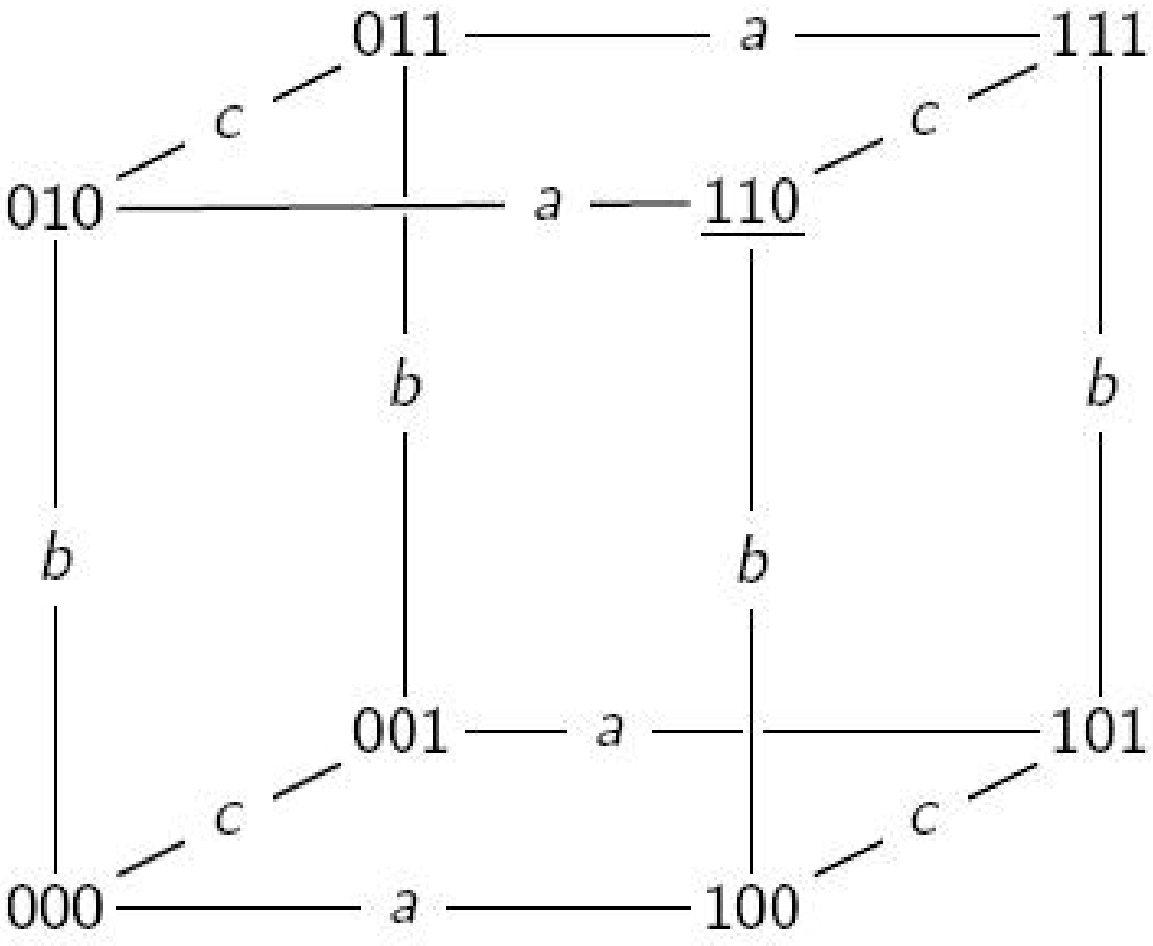
- Semantics: The effect of the public announcement of φ is the restriction of the epistemic state to all states where φ holds.
  'φ is the announcement' means 'φ is publicly and truthfully announced'.

After it is announced that p, everyone knows that p: $[p]Ep$
After it is announced that p, it is common knowledge that p: $[p]Cp$

The Muddy Children

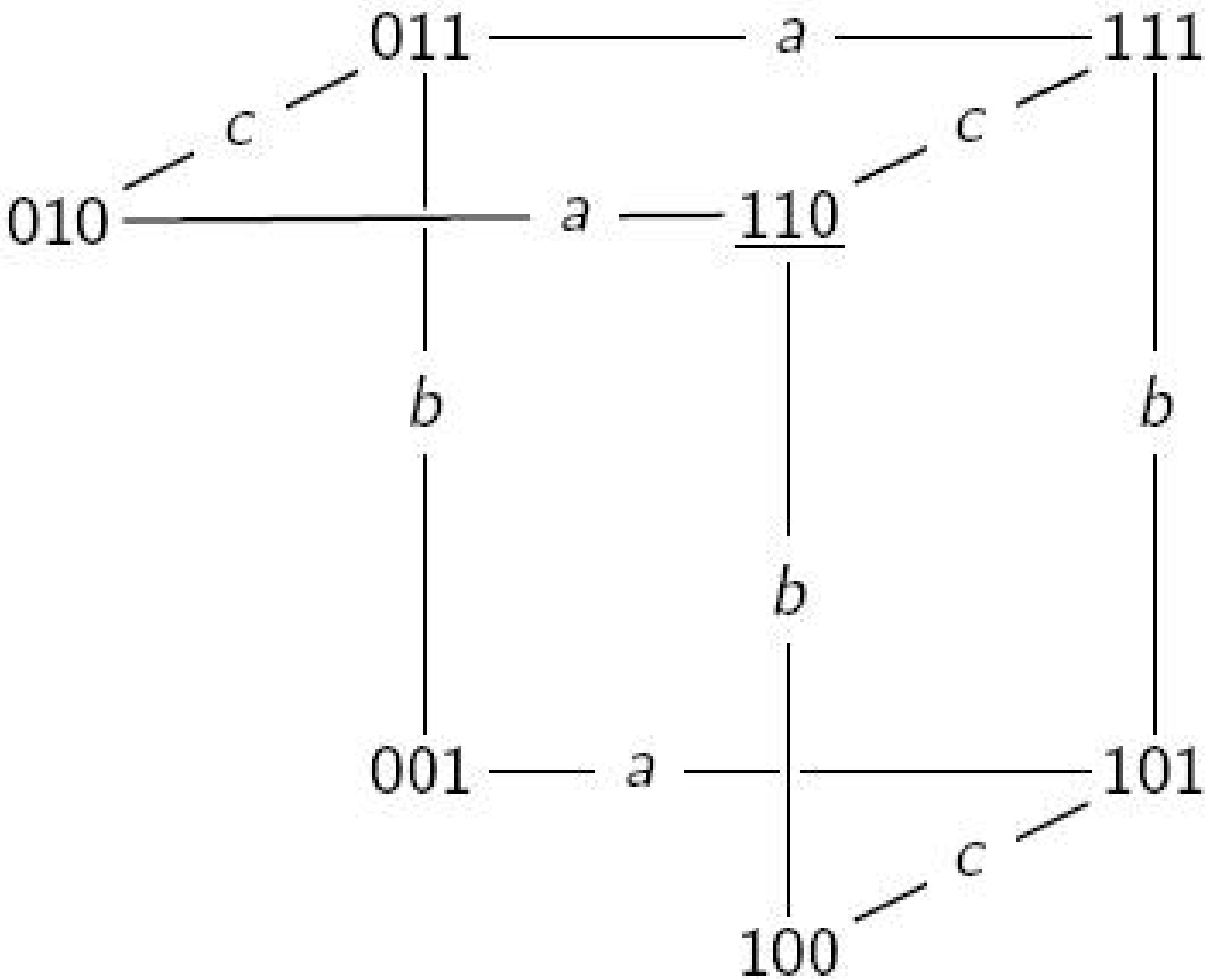
Picture by Marco Swaen

Epistemic Analysis



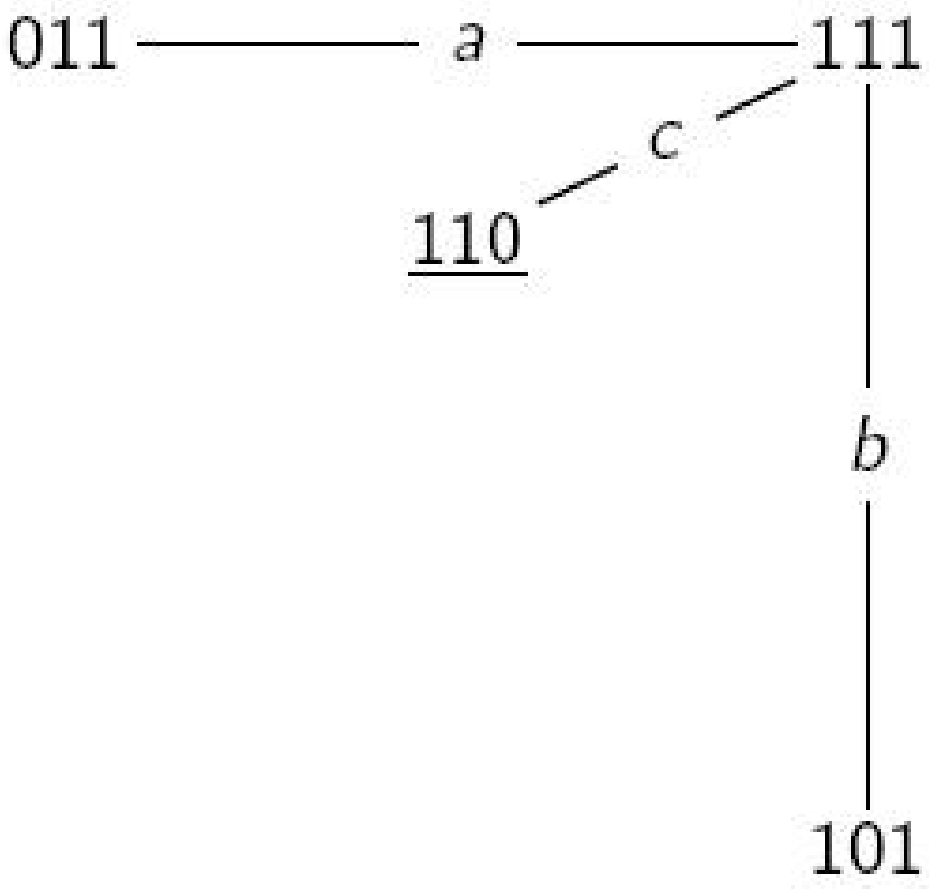
The children can see each other

Epistemic Analysis



At least one of you has mud on his or her forehead.

Epistemic Analysis



Will those who know whether they are muddy please step forward?

Epistemic Analysis

110

Will those who know whether they are muddy please step forward?

Outline

Limits on reasoning about others

- Many adults have difficulty in reasoning on higher orders than 2 without pen and paper:
  ”I do not know whether you know that Jan knows that I know that .....”
- Epistemic logic is an idealized model of human reasoning about knowledge, but it can still be a very useful tool
- Applications: network security and cryptography, study of social and coalitional interactions

Public Communication of Secrets: Russian Cards

From a pack of seven known cards $0, 1, 2, 3, 4, 5, 6$ Alice (a) and Bob (b) each draw three cards and Eve (c) gets the remaining card. How can Alice and Bob openly (publicly) inform each other about their cards, without Eve learning of any of their cards who holds it?

Suppose Alice draws $\{0, 1, 2\}$, Bob draws $\{3, 4, 5\}$, and Eve 6.

References

- Acknoledgement: slides are adapted from Jan van Eijck, Rineke Verbrugge