Dynamic Epistemic Logic

Hans van Ditmarsch
Logic, University of Sevilla, Spain, hvd@us.es

personal.us.es/hvd/

- multi-agent epistemic logic
- public announcement logic
- action model logic
- factual change, belief revision, temporal epistemic logic
- logic puzzles
- quantifying over propositions

Multi-agent Epistemic Logic

Multi-agent Epistemic Logic

Epistemic Logic

Anne draws one from a stack of three different cards 0, 1, and 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.
What does Anne know?

- 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.

Language

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

Descriptions of knowledge

- There is one agent Anne: $\{a\}$
- Propositional variables q_a for 'card q $(0, 1, 2)$ is held by Anne.'
- $K_a\varphi$ expresses 'Anne knows that φ'.
- $\hat{K}_a\varphi$ $(\neg K_a\neg\varphi)$ expresses 'Anne considers it possible that φ'.

- Anne holds card 0: 0_a
- Anne knows that she holds card 0: K_a0_a
- Anne does not know that card 1 is on the table: $\neg K_a1_t$
- Anne considers it possible that card 1 is not on the table: $\hat{K}_a\neg 1_t$
- 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: $K_a0_a \vee K_a1_a \vee K_a2_a$

Structures

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

- *domain* S is a nonempty set of states;
- R yields an *accessibility relation* $R_a \subseteq S \times S$ for every $a \in A$;
- *valuation* (function) $V : P \to \mathcal{P}(S)$.

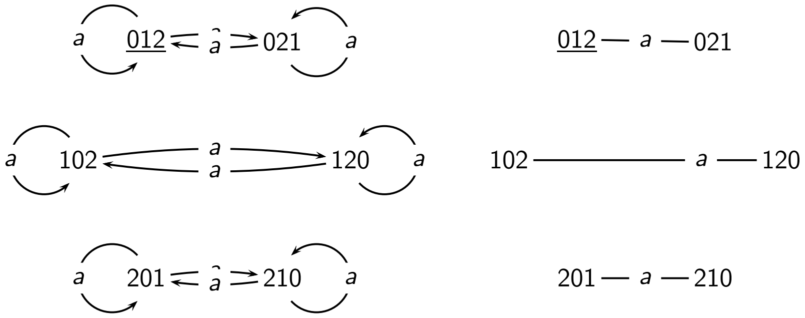
If all the relations R_a in M are equivalence relations, we call M an *epistemic model*. In that case, we write $\sim_a$ rather than R_a, and we represent the model as $M = \langle S, \sim, V \rangle$.

Epistemic state (M, s): epistemic model M with designated state s.

Example

$Hexa1 = \langle Hexa1, \sim, V \rangle$:

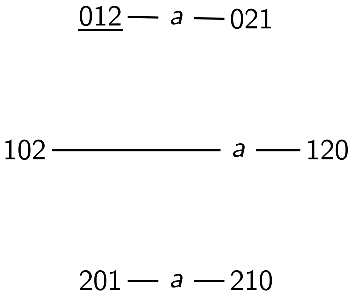
- $S = \{012, 021, 102, 120, 201, 210\}$
- $\sim_a = \{(012, 012), (012, 021), (021, 021), \ldots\}$
- $V(0_a) = \{012, 021\},\ V(1_a) = \{102, 120\},\ \ldots$



Truth

$$\begin{array}{lll}
M, s \models p & \text{iff} & s \in V(p) \\
M, s \models (\varphi \wedge \psi) & \text{iff} & M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models \neg\varphi & \text{iff} & \text{not } (M, s \models \varphi) \\
M, s \models K_a\varphi & \text{iff} & \text{for all } t \text{ such that } s \sim_a t \text{ it holds that } M, t \models \varphi
\end{array}$$

Example



$\underline{012}$ — a — 021

102 — a — 120

201 — a — 210

$Hexa1, 012 \models K_a 0_a$

$\Leftarrow$

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

$\Leftarrow$

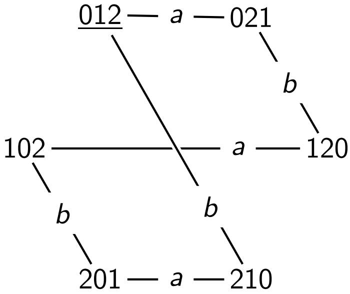
$Hexa1, 012 \models 0_a$ and $Hexa1, 021 \models 0_a$

$\Leftarrow$

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

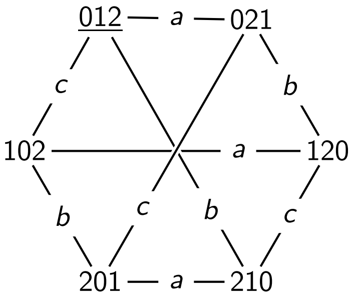
Two agents

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



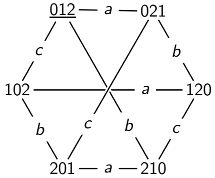
- Bill does not consider it possible that Anne has card 1: $\neg\hat{K}_b 1_a$
- Anne considers it possible that Bill considers it possible that she has card 1: $\hat{K}_a\hat{K}_b 1_a$
- Anne knows Bill to consider it possible that she has card 0: $K_a\hat{K}_b 0_a$

Three agents: Anne, Bill, Cath draw 0, 1, and 2



- Anne knows that Bill knows that Cath knows her own card: $K_a K_b (K_c 0_c \vee K_c 1_c \vee K_c 2_c)$
- Anne has card 0, but she considers it possible that Bill considers it possible that Cath knows that Anne does not have card 0: $0_a \wedge \hat{K}_a \hat{K}_b K_c \neg 0_a$

Example



$Hexa, 012 \models \hat{K}_a \hat{K}_b K_c \neg 0_a$

$\Leftarrow$

$012 \sim_a 021$ and $Hexa, 021 \models \hat{K}_b K_c \neg 0_a$

$\Leftarrow$

$021 \sim_b 120$ and $Hexa, 120 \models K_c \neg 0_a$

$\Leftarrow$

$\sim_c (120) = \{120, 210\}$, $Hexa, 120 \models \neg 0_a$ and $Hexa, 210 \models \neg 0_a$

$\Leftarrow$

$Hexa, 120 \not\models 0_a$ and $Hexa, 210 \not\models 0_a$

$\Leftarrow$

$120, 210 \notin V(0_a) = \{012, 021\}$

Properties of knowledge

- $K_a\varphi \to \varphi$ veridicality / truth axiom
- $K_a\varphi \to K_aK_a\varphi$ positive introspection
- $\neg K_a\varphi \to K_a\neg K_a\varphi$ negative introspection

Realistic assumptions for knowledge?

Negative introspection:
you are aware of everything you don't know. Really?
Weaker logic S4.2: $\hat{K}_aK_a\varphi \to K_a\hat{K}_a\varphi$ (.2, confluence)

Truth: everything you know is true. Really?
Weaker logic KD45 (introspective belief): $K_a\varphi \to \hat{K}_a\varphi$ (D, seriality)

Frame characterization

A Kripke model $\langle S, R, V \rangle$ *without* the valuation V of atoms is a Kripke *frame* $\langle S, R \rangle$. A formula φ is valid on a frame iff it is valid on all models based on that frame. Correspondence between Kripke frames and properties of knowledge:

- $K_a\varphi \rightarrow \varphi$ is valid on a frame, iff the frame is reflexive
- $K_a\varphi \rightarrow K_aK_a\varphi$ is valid on a frame, iff the frame is transitive
- $\neg K_a\varphi \rightarrow K_a\neg K_a\varphi$ is valid on a frame, iff it is euclidean

R_a is *euclidean*: if $R_a(s, s')$ and $R_a(s, s'')$, then $R_a(s', s'')$.

Frame characterization

We prove that the formula scheme $K\varphi \to \varphi$ is valid on a frame $F = \langle S, R \rangle$, if and only if R is reflexive.

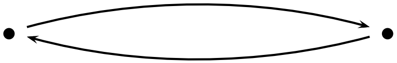
$\Leftarrow$ Let V be an arbitrary valuation on F. Consider the model $M = \langle F, V \rangle$. Suppose that $M, s \models K\varphi$. As R is reflexive, we have that $R(s,s)$. From $R(s,s)$ and $M, s \models K\varphi$ follows that $M, s \models \varphi$. So $M, s \models K\varphi \to \varphi$. As s and V were arbitrary, $F \models K\varphi \to \varphi$.

$\Rightarrow$ If R is not reflexive, there is a $s \in S$ such that *not* $R(s,s)$. Define, for some $p \in P$: $V(p) = S \setminus s$. Now, $M, s \models Kp$, because in all accessible worlds (which excludes s!) p is true. But $M, s \not\models p$. So $M, s \not\models Kp \to p$. But that means that the scheme $K\varphi \to \varphi$ is not valid on the frame F: on any not reflexive frame we have found a valuation, and state, and a formula, such that it is false.

Frame characterization

The correspondence does not work on the level of models.

Schema $K\varphi \to \varphi$ is valid on this non-reflexive model where a single atom *p* holds in both worlds:



Frame characterization

Axiom $K_a\varphi \to K_aK_a\varphi$ (4, positive introspection) corresponds to frame property $\forall s, t, u, R_a(s,t) \wedge R_a(t,u) \to R_a(s,u)$ (transitivity)

Proof. Take the dual $\hat{K}_a\hat{K}_a\varphi \to \hat{K}_a\varphi$ of the axiom.

$\Leftarrow$ Let $M, s \models \hat{K}_a\hat{K}_a\varphi$. Then there are t, u with $R_a(s,t)$ and $R_a(t,u)$ such that $M, u \models \varphi$. From $R_a(s,t)$, $R_a(t,u)$ and transitivity follows $R_a(s,u)$. From $R_a(s,u)$ and $M, u \models \varphi$ follows $M, s \models \hat{K}_a\varphi$.

$\Rightarrow$ Given a non-transitive frame. There must be s, t, u with $R_a(s,t)$ and $R_a(t,u)$ but not $R_a(s,u)$. Consider model M such that atom p only true at u. We now have $M, s \models \hat{K}_a\hat{K}_a p$ but $M, s \not\models \hat{K}_a p$. Therefore $\hat{K}_a\hat{K}_a\varphi \to \hat{K}_a\varphi$ does not hold on that frame for all φ and for all valuations.

Multi-agent frame characterization

Multi-agent frame property: $K_a\varphi \rightarrow K_b\varphi$.
Corresponds to $R_b \subseteq R_a$.
(Agent b knows more than agent a?)

Axiomatization

all instantiations of propositional tautologies

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

$K_a\varphi \to \varphi$

$K_a\varphi \to K_aK_a\varphi$

$\neg K_a\varphi \to K_a\neg K_a\varphi$

From φ and $\varphi \to \psi$, infer ψ

From φ, infer $K_a\varphi$

Give derivations of:

$K_ap \to K_a(p \vee q)$

$K_a\neg p \to \neg K_ap$

History

- von Wright 1951: An Essay in Modal Logic
- Hintikka 1962: Knowledge and Belief
- Fagin, Halpern, Moses and Vardi 1995: Reasoning about Knowledge
- Meyer and van der Hoek 1995: Epistemic Logic for AI and Computer Science

General knowledge and common knowledge

You forgot if you already passed the Channel Tunnel... When driving on a one-lane road, will you swerve to the left or to the right when other traffic approaches? How do you know that the other car knows that one is to drive on the left?

You are celebrating Sinterklaas (St. Nicholas) with family friends. How will you behave if its generally known that your 8-year old niece does not believe in Sinterklaas? And if it is common knowledge?

General knowledge and common knowledge

General knowledge:
$E_G\varphi := K_1\varphi \wedge K_2\varphi \wedge ... \wedge K_{\text{last}}\varphi$

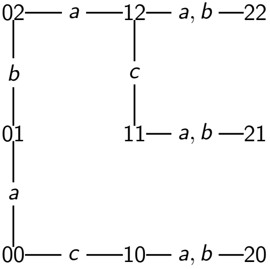
Common knowledge:
$C_G\varphi := \varphi \wedge E_G\varphi \wedge E_G E_G\varphi \wedge ...$
or
$C_G\varphi := \varphi \wedge K_1\varphi \wedge K_2\varphi \wedge K_1 K_1\varphi \wedge K_1 K_2\varphi \wedge \ldots K_1 K_1 K_1\varphi \ldots$

$C_G\varphi \leftrightarrow \varphi \wedge E_G C_G\varphi$

Computing transitive closure

$$\sim_B := \Big(\bigcup_{a\in B} \sim_a\Big)^*$$

R^* is the transitive and reflexive closure of a binary relation R: points s and t are R^*-related, if there is a path (of length 0 or more) of R-links between them.



What is the partition on these nine states for a?
For group $\{a, b\}$? For group $\{a, c\}$? For group $\{a, b, c\}$?

Epistemic Logic with Common Knowledge

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

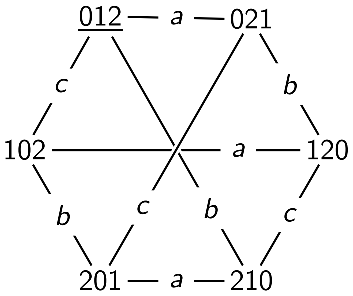
$$\sim_B := (\bigcup_{a \in B} \sim_a)^*$$

$M, s \models C_B\varphi$ iff for all t : $s \sim_B t$ implies $M, t \models \varphi$

Common knowledge has the properties of knowledge:

- $C_B\varphi \rightarrow \varphi$ veridicality / truth axiom
- $C_B\varphi \rightarrow C_BC_B\varphi$ positive introspection
- $\neg C_B\varphi \rightarrow C_B\neg C_B\varphi$ negative introspection

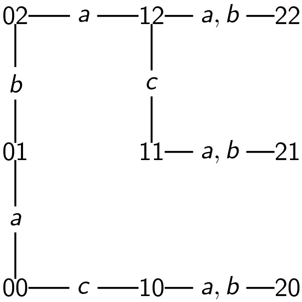
Example



$Hexa, 012 \models C_{abc}(K_a 0_a \vee K_a 1_a \vee K_a 2_a)$
(it is public knowledge that Anne knows her card)

$Hexa \models C_{ab}\varphi \rightarrow C_{bc}\varphi$
(a and b share the same knowledge as b and c)

Example



Which of the following are true / false:

$$
\begin{aligned}
11 &\models K_c(x = 1) \\
11 &\models C_{ac}(y \neq 0) \\
10 &\models C_{ab}(x \geq 1) \\
02 &\models C_{ab}((y = 2) \rightarrow C_{cb}(x > 0))
\end{aligned}
$$

Axiomatization

$C_B(\varphi \to \psi) \to (C_B\varphi \to C_B\psi)$
$C_B\varphi \to (\varphi \wedge E_B C_B\varphi)$
$C_B(\varphi \to E_B\varphi) \to (\varphi \to C_B\varphi)$
From φ, infer $C_B\varphi$

Give derivations of:
$C_B\varphi \to C_B C_B\varphi$
$\neg C_B\varphi \to C_B\neg C_B\varphi$

Variations:

- reflexive and transitive closure, or just transitive closure?
- just transitive: $C_B(\varphi \to E_B\varphi) \to (E_B\varphi \to C_B\varphi)$
- axiom or derivation rule: From $\varphi \to E_B\varphi$, infer $\varphi \to C_B\varphi$

Common knowledge and common belief

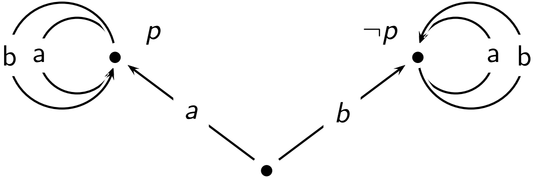
Common knowledge has the properties of knowledge.

- $C_B\varphi \to \varphi$
- $C_B\varphi \to C_B C_B\varphi$
- $\neg C_B\varphi \to C_B\neg C_B\varphi$

Common belief does not have (all) the properties of belief.

- $C_B\varphi \to \neg C_B\neg\varphi$
- $C_B\varphi \to C_B C_B\varphi$
- not valid is: $\neg C_B\varphi \to C_B\neg C_B\varphi$ (negative introspection)

Countermodel (for which $\neg C_{ab}p$, but $\hat{K}_a C_{ab}p$):



Relativized common knowledge

Common knowledge was defined as:

$$M, s \models C_B\varphi \quad \text{iff} \quad \text{for all } t, s \sim_B t \text{ implies } M, t \models \varphi$$

Consider the novel construct $C_B^\psi \varphi$ for

'along all the B-paths satisfying ψ it holds that φ.'

This is called common knowledge of φ *relativized* to ψ.

Let $s \sim_a^\psi t$ iff $s \sim_a t$ and $M, t \models \psi$, and $\sim_B^\psi := (\bigcup_{a \in B} \sim_a^\psi)^+$.
Then we define:

$$M, s \models C_B^\psi\varphi \quad \text{iff} \quad \text{for all } t, s \sim_B^\psi t \text{ implies } M, t \models \varphi$$

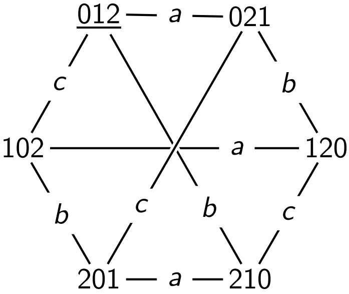
We have that $C_B^\top \varphi$ iff $C_B\varphi$. Epistemic logic with relativized common knowledge is more expressive than epistemic logic.

Distributed knowledge

Construct $D_B\varphi$ for "it is distributed knowledge among B that φ".

$$M, s \models D_B\varphi \quad \text{iff} \quad s \sim_a t \text{ for all } a \in B, \text{ implies } M, t \models \varphi$$

E.g., agents a, b, c have distributed knowledge of the card deal.



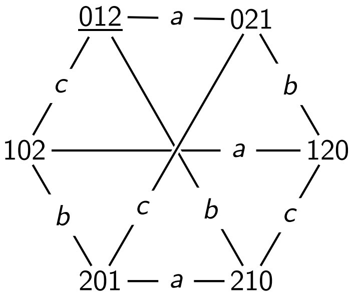
History

- Lewis 1969: Convention
- Friedell 1969: On the structure of shared awareness
- Aumann 1976: Agreeing to disagree
- Barwise 1988: Three views of common knowledge

Public announcements

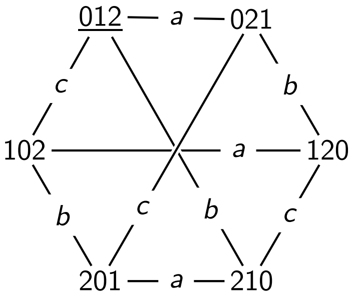
Public announcements

Example



- After Anne says that she does not have card 1, Cath knows that Bill has card 1.
- After Anne says that she does not have card 1, Cath knows Anne's card.
- Bill still doesn't know Anne's card after that.

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 Announcement Logic: language

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

Write $\langle\varphi\rangle\psi$ for $\neg[\varphi]\neg\psi$

For $[\varphi]\psi$ read "after the announcement of φ, ψ (is true)."

For $\langle\varphi\rangle\psi$ read "φ is true and after the announcement of φ, ψ."

Public Announcement Logic: semantics

The effect of the public announcement of φ is the restriction of the epistemic state to all states where φ holds. So, 'announce φ' can be seen as an epistemic state transformer, with a corresponding dynamic modal operator $[\varphi]$.

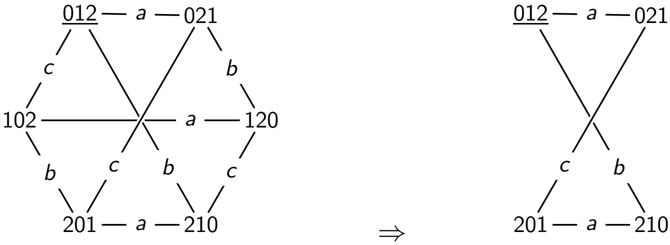
'φ is the announcement'
means
'φ is publicly and truthfully announced'.

$$M, s \models [\varphi]\psi \text{ iff } (M, s \models \varphi \text{ implies } M|\varphi, s \models \psi)$$

$M|\varphi := \langle S', \sim', V' \rangle$:

$$\begin{array}{lcl} S' & := & [\![\varphi]\!]_M \quad := \{s \in S \mid M, s \models \varphi\} \\ \sim'_a & := & \sim_a \cap ([\![\varphi]\!]_M \times [\![\varphi]\!]_M) \\ V'(p) & := & V(p) \cap [\![\varphi]\!]_M \end{array}$$

Example announcement in Hexa



$Hexa, 012 \models \langle \neg 1_a \rangle K_c 0_a$

$\Leftarrow$

$Hexa, 012 \models \neg 1_a$ and $Hexa|\neg 1_a, 012 \models K_c 0_a$

$\Leftarrow$

$Hexa, 012 \models \neg 1_a$ and $(Hexa|\neg 1_a, 012 \models 0_a$ and $\sim_c(012) = \{012\})$

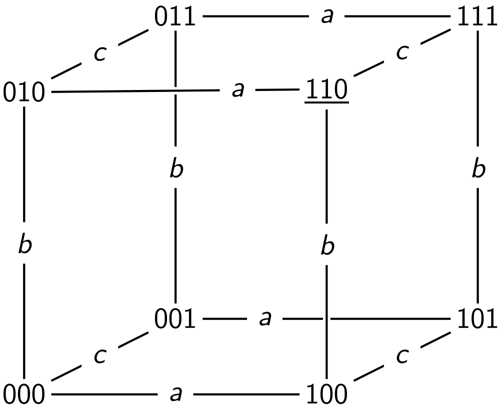
$\Leftarrow$

$012 \neq V(1_a)$ and $012 \in V'(0_a)$

Muddy Children

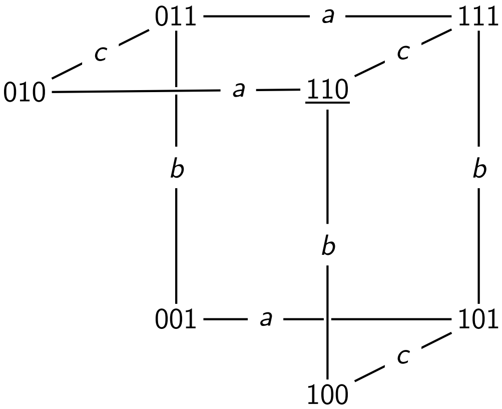
A group of children has been playing outside and are called back into the house by their father. The children gather round him. As one may imagine, some of them have become dirty from the play and in particular: they may have mud on their forehead. Children can only see whether other children are muddy, and not if there is any mud on their own forehead. All this is commonly known, and the children are, obviously, perfect logicians. Father now says: “At least one of you has mud on his or her forehead.” And then: “Will those who know whether they are muddy please step forward.” If nobody steps forward, father keeps repeating the request. What happens?

Muddy Children



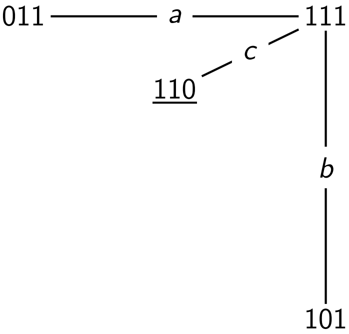
Given: The children can see each other

Muddy Children



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

Muddy Children



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

Muddy Children

$\underline{110}$

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

Muddy Children

German translation of Rabelais' Gargantua and Pantagruel: Gottlob Regis, *Meister Franz Rabelais der Arzeney Doctoren Gargantua und Pantagruel, usw.*, Barth, Leipzig, 1832.

Ungelacht pfetz ich dich. Gesellschaftsspiel. Jeder zwickt seinen rechten Nachbar an Kinn oder Nase; wenn er lacht, giebt er ein Pfand. Zwei von der Gesellschaft sind nämlich im Complot und haben einen verkohlten Korkstöpsel, woran sie sich die Finger, und mithin denen, die sie zupfen, die Gesichter schwärzen. Diese werden nun um so lächerlicher, weil jeder glaubt, man lache über den anderen.

I pinch you without laughing. Parlour game. Everybody pinches his right neighbour into chin or nose; if one laughs, one must give a pledge. Two in the round have secretly blackened their fingers on a charred piece of cork, and hence will blacken the faces of their neighbours. These neighbours make a fool of themselves, since they both think that everybody is laughing about the other one.

Axiomatization

$[\varphi]p \leftrightarrow (\varphi \to p)$
$[\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[\varphi]\psi)$
$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$
$[\varphi]K_a\psi \leftrightarrow (\varphi \to K_a[\varphi]\psi)$
$[\varphi][\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi$
From φ, infer $[\psi]\varphi$
From $\chi \to [\varphi]\psi$ and $\chi \wedge \varphi \to E_B\chi$, infer $\chi \to [\varphi]C_B\psi$

Expressivity (Plaza, Gerbrandy):
Every formula in the language of public announcement logic without common knowledge is equivalent to a formula in the language of epistemic logic.

Sequence of announcements

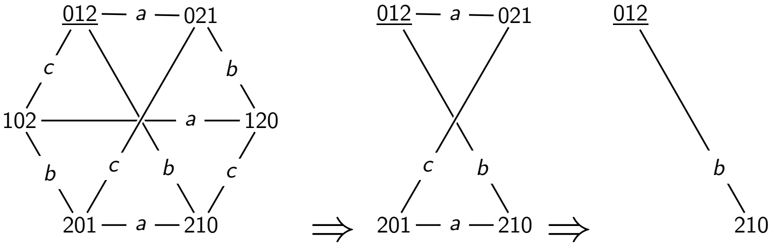
$$[\varphi][\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi$$

Anne does not have card 1, and Cath now knows Anne's card.
Sequence of two announcements:

$$\neg 1_a \; ; \; (K_c 0_a \vee K_c 1_a \vee K_c 2_a)$$

Single announcement:

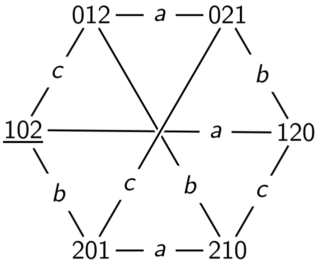
$$\neg 1_a \wedge [\neg 1_a](K_c 0_a \vee K_c 1_a \vee K_c 2_a)$$



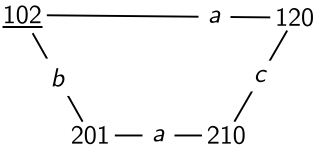
Announcement and knowledge

$$[\varphi]K_a\psi \leftrightarrow (\varphi \rightarrow K_a[\varphi]\psi)$$

$Hexa, 012 \models [\neg 0_a]K_c 0_a$
$Hexa,\ 012 \not\models K_c[\neg 0_a]0_a$
$Hexa,\ 012 \models \neg 0_a \rightarrow K_c[\neg 0_a]0_a$

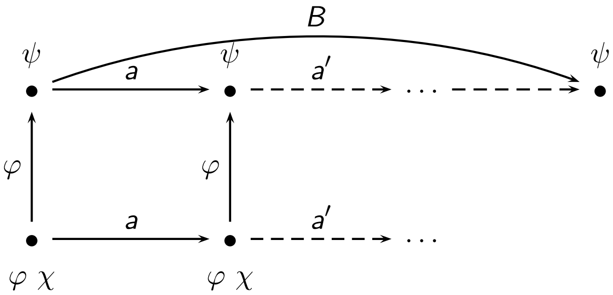


$\Longrightarrow \neg 0_a$



Announcement and common knowledge

From $\chi \rightarrow [\varphi]\psi$ and $\chi \wedge \varphi \rightarrow E_B\chi$, infer $\chi \rightarrow [\varphi]C_B\psi$



'Common knowledge induction' is a special case.
Take $\varphi := \top$ and $\chi := \psi$:
$C_B(\psi \rightarrow E_B\psi) \rightarrow (\psi \rightarrow C_B\psi)$

Relativized common knowledge

Announcement and common knowledge

From $\chi \rightarrow [\varphi]\psi$ and $\chi \wedge \varphi \rightarrow E_B\chi$, infer $\chi \rightarrow [\varphi]C_B\psi$

Announcement and relativized common knowledge

$[\varphi]C_B^{\chi}\psi \leftrightarrow C_B^{\varphi \wedge [\varphi]\chi}[\varphi]\psi$

Derived principle for common knowledge:

$[\varphi]C_B\psi \leftrightarrow C_B^{\varphi}[\varphi]\psi$

Unsuccessful updates

Postulate of success:

$$\varphi \to \langle \varphi \rangle C_A \varphi$$

Announcement of a *fact* always makes it public:

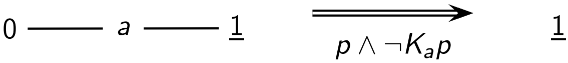
$$\models [p] C_A p$$

Announcements of non-facts do not have to make them public:

$$\not\models [\varphi] C_A \varphi$$

It can be even worse:

$$\models [p \land \neg K_a p] \neg (p \land \neg K_a p)$$



Unsuccessful updates

Successful formulas: $[\varphi]\varphi$ is valid.
Because $[\varphi]\varphi$ iff $[\varphi]C_A\varphi$ iff $\varphi \to [\varphi]C_A\varphi$

Which formulas are successful?

- $C_A\varphi$, for any φ in the language (but *only* public knowledge)
- the language fragment of positive formulas
  $\varphi ::= p|\neg p|\varphi \vee \varphi|\varphi \wedge \varphi|K_a\varphi|[\neg\varphi]\varphi$.
- the formula $\neg Kp$... (Lei Xian)

Characterization of one-agent successful formulas:
Holliday & Icard AiML 2010
Characterization of multi-agent successful formulas:
unknown!

Unsuccessful updates

At least I cannot learn from my own announcements...

So ignorance may become knowledge,
but at least knowledge may not become ignorance...

Unsuccessful updates

At least I cannot learn from my own announcements...

So ignorance may become knowledge,
but at least knowledge may not become ignorance...

Wrong again, same example...
Add an agent i with identity access on the model ('the observer').
After agent i announces $K_i(p \land \neg K_a p)$, this formula is false.
Agent i becomes ignorant (about that) from her own announcement.
(E.g.) Agent i becomes knowledgeable about $K_a p$!

$$0 \;\text{———}\; a \;\text{———}\; \underline{1} \qquad \xRightarrow[K_i(p \land \neg K_a p)]{} \qquad \underline{1}$$

Alternative semantics for public announcement

Truthful announcements

$$M, s \models [\varphi]\psi \ \text{ iff } \ (M, s \models \varphi \text{ implies } M|\varphi, s \models \psi)$$

$M|\varphi := \langle S', \sim', V' \rangle$:

$$\begin{array}{lcl} S' & := & [\![\varphi]\!]_M \\ \sim'_a & := & \sim_a \cap ([\![\varphi]\!]_M \times [\![\varphi]\!]_M) \\ V'(p) & := & V(p) \cap [\![\varphi]\!]_M \end{array}$$

Believed announcements

$$M, s \models [\varphi]\psi \ \text{ iff } \ M^\varphi, s \models \psi$$

$M^\varphi := \langle S, \sim'', V \rangle$:

$$\sim''_a \ := \ \sim_a \cap (S \times [\![\varphi]\!]_M)$$

Remove arrows to states where the announcement is false.

Now announce something believed to be false...

History

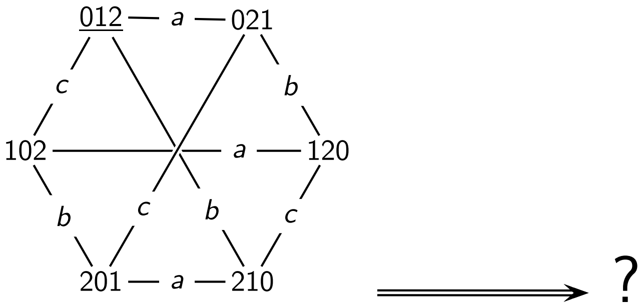
- Plaza 1989: Logics of Public Communications
- Gerbrandy & Groeneveld 1997: Reasoning about Information Change
- Baltag, Moss & Solecki 1998: The Logic of Common Knowledge, Public Announcements, and Private Suspicions

Action models

Action models

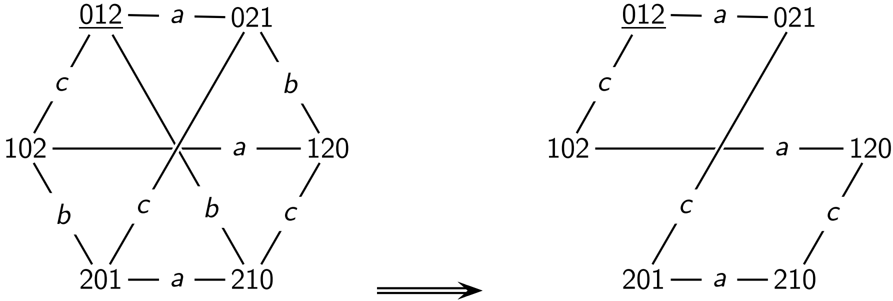
What we cannot do yet...

(Anne holds 0, Bill holds 1, and Cath holds 2.) Anne shows (only) Bill her card. (She shows card 0.) Cath cannot see the face of the shown card, but notices that a card is being shown.



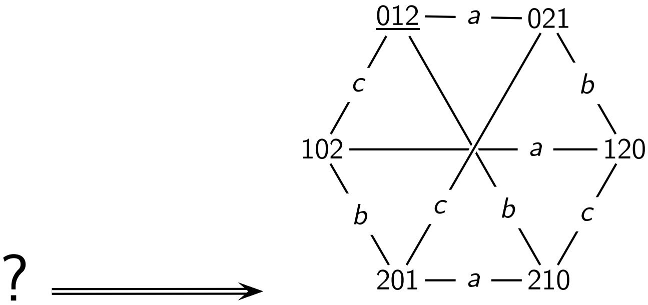
What we cannot do yet...

(Anne holds 0, Bill holds 1, and Cath holds 2.) Anne shows (only) Bill her card. (She shows card 0.) Cath cannot see the face of the shown card, but notices that a card is being shown.



What we also cannot do yet...

Anne holds 0, Bill holds 1, and Cath holds 2. Players only know their own cards.



Epistemic modeling

- Given is an informal description of a situation
- The modeler tries to determine:
  - The set of relevant propositions
  - The set of relevant agents
  - The set of states
  - An indistinguishability relation over these worlds for each agent

Dynamic modeling

- Given is an informal description of a situation and an event that takes place in that situation.
- The modeler first models the epistemic situation, and then tries to determine:
  - The set of possible events
  - The preconditions for the events
  - An indistinguishability relation over these events for each agent

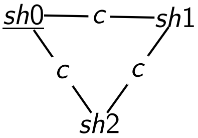
Action models

An action model M is a structure $\langle \mathsf{S}, \sim, \mathsf{pre} \rangle$

- S is a *finite* domain of action points or events
- $\sim_a$ is an equivalence relation on S
- $\mathsf{pre} : \mathsf{S} \to \mathcal{L}$ is a precondition function that assigns a precondition to each $\mathsf{s} \in \mathsf{S}$.

Showing a card

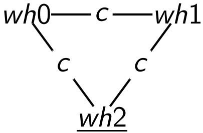
(Anne holds 0, Bill holds 1, and Cath holds 2.) Anne shows (only) Bill card 0. Cath cannot see the face of the shown card, but notices that a card is being shown.



- $S = \{sh0, sh1, sh2\}$
- $\sim_a = \{(s, s) \mid s \in S\}$
- $\sim_b = \{(s, s) \mid s \in S\}$
- $\sim_c = S \times S$
- $pre(sh0) = 0_a$
- $pre(sh1) = 1_a$
- $pre(sh2) = 2_a$

Whispering

Bill asks Anne to tell him a card that she doesn't have. Anne whispers in Bill's ear "I don't have card 2". Cath notices that the question is answered, but cannot hear the answer.



- $\mathsf{S} = \{\mathsf{wh0}, \mathsf{wh1}, \mathsf{wh2}\}$
- $\sim_a = \{(\mathsf{s}, \mathsf{s}) \mid \mathsf{s} \in \mathsf{S}\}$
- $\sim_b = \{(\mathsf{s}, \mathsf{s}) \mid \mathsf{s} \in \mathsf{S}\}$
- $\sim_c = \mathsf{S} \times \mathsf{S}$
- $\mathsf{pre}(\mathsf{wh0}) = \neg 0_a$
- $\mathsf{pre}(\mathsf{wh1}) = \neg 1_a$
- $\mathsf{pre}(\mathsf{wh2}) = \neg 2_a$

What do you learn from an action?

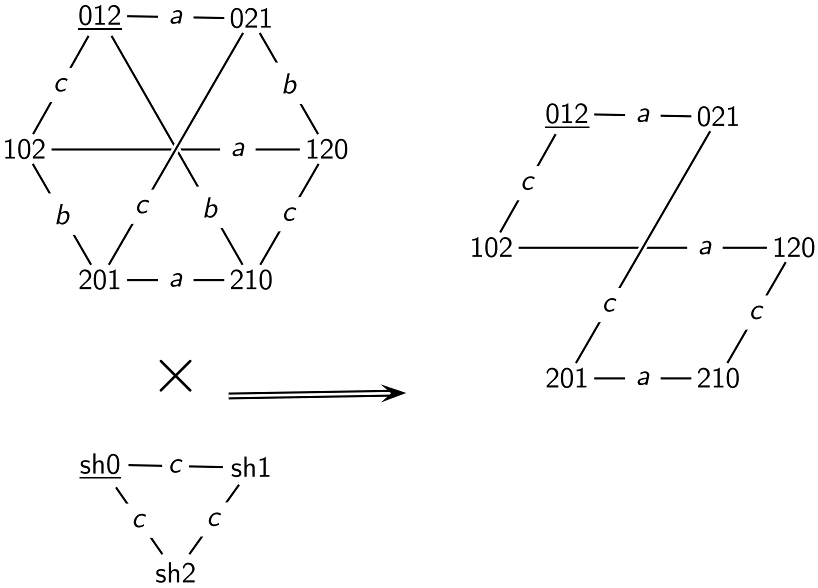
- Firstly, if you can distinguish two actions, then you can also distinguish the states that result from executing the action.
- Secondly, you do not forget anything due to an action. States that you could distinguish before an action are still distinguishable.

Product update

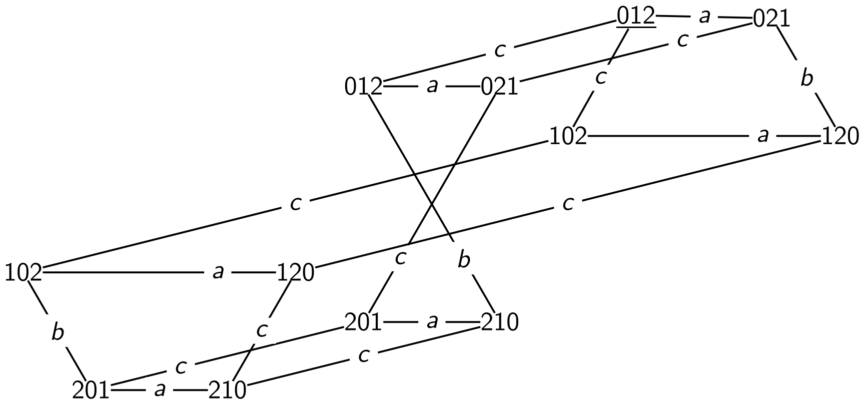
Given are an epistemic state (M, s) with $M = \langle S, \sim, V \rangle$ and an action model (M, s) with $\mathsf{M} = \langle \mathsf{S}, \sim, \mathsf{pre} \rangle$. The result of executing (M, s) in (M, s) is $(M \otimes \mathsf{M}, (s, \mathsf{s}))$ where $M \otimes \mathsf{M} = \langle S', \sim', V' \rangle$ such that:

- $S' = \{(s, \mathsf{s}) \mid s \in S, \mathsf{s} \in \mathsf{S}, \text{ and } M, s \models \mathsf{pre}(\mathsf{s})\}$
- $(s, \mathsf{s}) \sim'_a (t, \mathsf{t})$ iff $(s \sim_a t \text{ and } \mathsf{s} \sim_a \mathsf{t})$
- $(s, \mathsf{s}) \in V'(p)$ iff $s \in V(p)$

Anne shows card 0 to Bill



Anne whispers ‘not 0’ to Bill



Language

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid C_B\varphi \mid [\mathsf{M}, \mathsf{s}]\varphi$$

Semantics

$$
\begin{array}{lll}
M, s \models p & \text{:iff} & s \in V(p) \\
M, s \models \neg\varphi & \text{:iff} & M, s \not\models \varphi \\
M, s \models \varphi \wedge \psi & \text{:iff} & M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models K_a\varphi & \text{:iff} & \text{for all } s' \in S : s \sim_a s' \text{ implies } M, s' \models \varphi \\
M, s \models C_B\varphi & \text{:iff} & \text{for all } s' \in S : s \sim_B s' \text{ implies } M, s' \models \varphi \\
M, s \models [\mathsf{M}, \mathsf{s}]\varphi & \text{:iff} & M, s \models \mathsf{pre}(\mathsf{s}) \text{ implies } \; M \otimes \mathsf{M}, (s, \mathsf{s}) \models \varphi
\end{array}
$$

Syntax and semantics

- Are syntax and semantics clearly separated?

YES

Axiomatization

$[\mathrm{M}, \mathrm{s}]p \leftrightarrow (\mathrm{pre}(\mathrm{s}) \rightarrow p)$

$[\mathrm{M}, \mathrm{s}]\neg\varphi \leftrightarrow (\mathrm{pre}(\mathrm{s}) \rightarrow \neg[\mathrm{M}, \mathrm{s}]\varphi)$

$[\mathrm{M}, \mathrm{s}](\varphi \wedge \psi) \leftrightarrow ([\mathrm{M}, \mathrm{s}]\varphi \wedge [\mathrm{M}, \mathrm{s}]\psi)$

$[\mathrm{M}, \mathrm{s}]K_a\varphi \leftrightarrow (\mathrm{pre}(\mathrm{s}) \rightarrow \bigwedge_{\mathrm{s} \sim_a \mathrm{t}} K_a[\mathrm{M}, \mathrm{t}]\varphi)$

$[\mathrm{M}, \mathrm{s}][\mathrm{M}', \mathrm{s}']\varphi \leftrightarrow [(\mathrm{M}, \mathrm{s}); (\mathrm{M}', \mathrm{s}')]\varphi$

From φ, infer $[\mathrm{M}, \mathrm{s}]\varphi$

Let (M, s) be an action model and let a set of formulas χ_t for every t such that $\mathrm{s} \sim_B \mathrm{t}$ be given. From $\chi_\mathrm{t} \rightarrow [\mathrm{M}, \mathrm{t}]\varphi$ and $(\chi_\mathrm{t} \wedge \mathrm{pre}(\mathrm{t})) \rightarrow K_a\chi_\mathrm{u}$ for every $\mathrm{t} \in \mathrm{S}$ such that $\mathrm{s} \sim_B \mathrm{t}$, $a \in B$ and $\mathrm{t} \sim_a \mathrm{u}$, infer $\chi_\mathrm{s} \rightarrow [\mathrm{M}, \mathrm{s}]C_B\varphi$.

Every formula in the language of action model logic without common knowledge is equivalent to a formula in the language of epistemic logic.

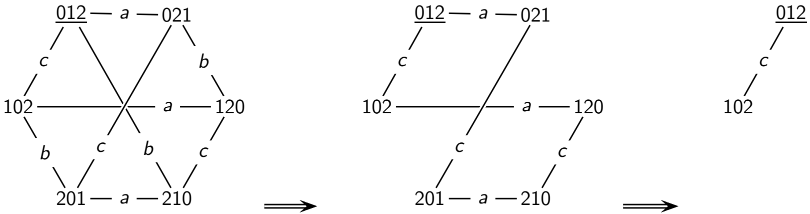
Composition of action models

Given action models (M, s) with $\mathsf{M} = \langle \mathsf{S}, \sim, \mathsf{pre} \rangle$ and $(\mathsf{M}', \mathsf{s}')$ with $\mathsf{M}' = \langle \mathsf{S}', \sim', \mathsf{pre}' \rangle$, their composition is the action model $(\mathsf{M}; \mathsf{M}', (\mathsf{s}, \mathsf{s}'))$ with $\mathsf{M}; \mathsf{M}' = \langle \mathsf{S}'', \sim'', \mathsf{pre}'' \rangle$:

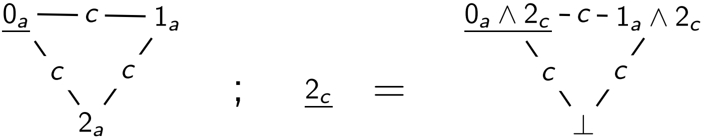
- $\mathsf{S}'' = \{(\mathsf{s}, \mathsf{s}') \mid \mathsf{s} \in \mathsf{S}, \mathsf{s}'\} \in \mathsf{S}'$
- $(\mathsf{s}, \mathsf{s}') \sim''_a (\mathsf{t}, \mathsf{t}')$ iff $(\mathsf{s} \sim_a \mathsf{t}$ and $\mathsf{s} \sim_a \mathsf{t})$
- $\mathsf{pre}''(\mathsf{s}, \mathsf{s}') = \langle \mathsf{M}, \mathsf{s} \rangle \mathsf{pre}'(\mathsf{s}')$

Action model composition – example

Anne shows 0 to Bill, after which Cath announces that she has 2.



Composition of action models

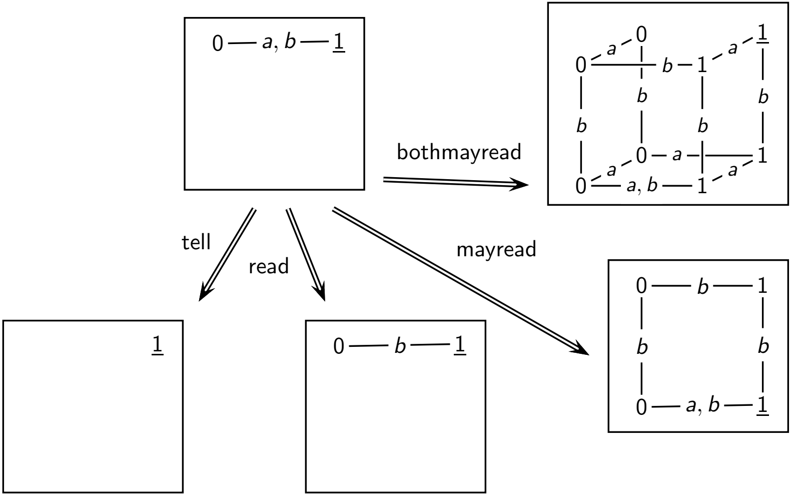


Other example: reading a letter

Anne and Bill are sitting at a table. A messenger comes in and delivers a letter to Anne. On the cover is written "urgently requested data on United Agents."

- tell
  Anne reads the letter aloud. United Agents is doing well.
- read
  Bill sees that Anne reads the letter. (United Agents is doing well.)
- mayread
  Bill leaves the table and orders a drink at the bar so that Anne may have read the letter while he was away. (She does not; United Agents is doing well.)
- bothmayread
  Both may have read the letter. (Both read the letter; United Agents is doing well.)

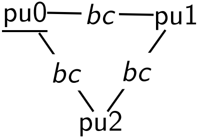
Other example: reading a letter

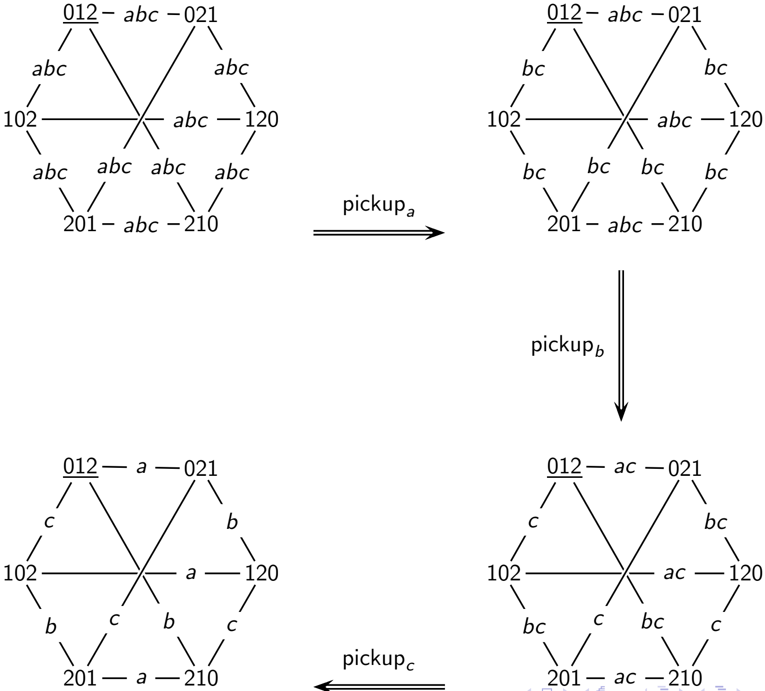


Closing example: picking up cards

Three players Anne, Bill, Cath are each dealt one of cards $0, 1, 2$.

- pickup$_a$: Anne picks up her card and looks at it. It is card 0.
- pickup$_b$: Bill picks up his card and looks at it. It is card 1.
- pickup$_c$: Cath picks up her card and looks at it. It is card 2.




012
021
102
120
201
210
abc
pickup$_a$
bc
abc
pickup$_b$
c
bc
ac
pickup$_c$
a
b
c

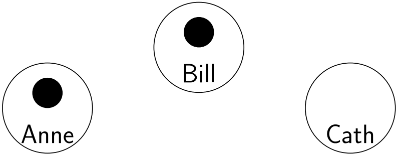
History

- Baltag, Moss & Solecki 1998: The Logic of Common Knowledge, Public Announcements, and Private Suspicions
- van Ditmarsch, van der Hoek & Kooi 2007: Dynamic Epistemic Logic
- van Benthem, van Eijck, Kooi 2006: Logics of communication and change

Further developments in dynamic epistemic logic

- Factual change
- Belief revision
- Temporal epistemic logic

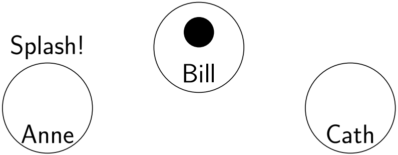
Factual change — Muddy Children again



There are three children, Anne, Bill, and Cath. Anne and Bill have mud on their foreheads. Father announces:

- At least one of you is muddy.
- If you know whether you are muddy, step forward. (Nobody steps forward.)
- If you know whether you are muddy, step forward. (Anne and Bill step forward.)

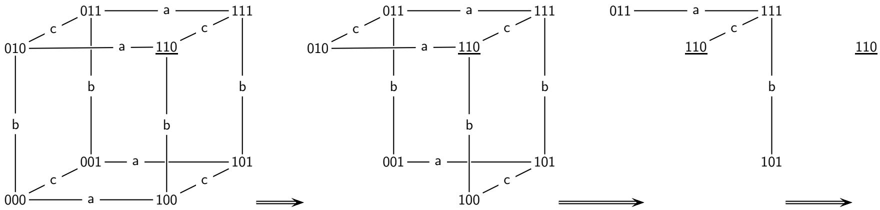
Cleaning Muddy Children



There are three children, Anne, Bill, and Cath. Anne and Bill have mud on their foreheads. Father announces:

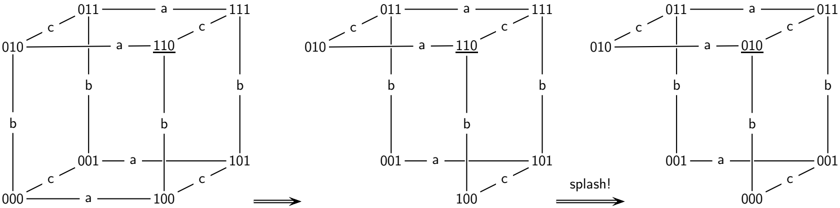
- At least one of you is muddy.
- **Splash!** *Father empties a bucket of water over Anne.*
- If you know whether you are muddy, step forward. (...?)
- If you know whether you are muddy, step forward. (...?)

Standard: Anne and Bill are muddy



- At least one child is muddy.
- Nobody steps forward.
- Anne and Bill step forward.

Non-standard: Anne and Bill are muddy, Anne is cleaned



- At least one child is muddy.
- *Father empties a bucket of water over Anne* (splash!)
- If you know whether you are muddy, step forward. (...?)
- If you know whether you are muddy, step forward. (...?)

Public factual change

Language

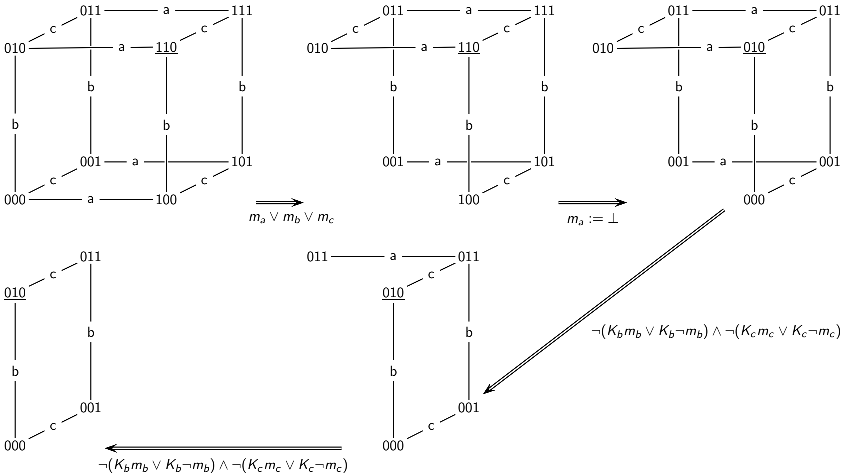
$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_a\varphi \mid C_A\varphi \mid [\varphi]\psi \mid [p := \varphi]\psi$$

Semantics

$$M, s \models [p := \varphi]\psi \quad \text{iff} \quad M_{p:=\varphi}, s \models \psi$$

$M_{p:=\varphi}$ is as M except that $V(p) = [\![\varphi]\!]_M$.

reduction principle: $[p := \varphi]p \leftrightarrow \varphi$.



At father's second request, Cath learns that Anne knows that she **was** *initially dirty*

Factual change with action models

An action model M is a structure $\langle \mathsf{S}, \sim, \mathsf{pre}, \mathsf{post}\rangle$

- S is a *finite* domain of action points or events
- $\sim_a$ is an equivalence relation on S
- $\mathsf{pre} : \mathsf{S} \rightarrow \mathcal{L}$ is a precondition function that assigns a precondition to each $\mathsf{s} \in \mathsf{S}$.
- $\mathsf{post} : \mathsf{S} \rightarrow (P \rightarrow \mathcal{L})$ assigns a postcondition to each action point for each atom (finitely different from the identity).

For s with $\mathsf{pre}(\mathsf{s}) = \varphi$ and $\mathsf{post}(\mathsf{s})(p_1) = \psi_1, \dots$ write:

In case of the event s: if φ, then $p_1 := \psi_1$, ..., and $p_n := \psi_n$.

Execution of action models with factual change. Same, except:

- $V'(p) = \{(s, \mathsf{s}) \mid (M, s) \models \mathsf{post}(\mathsf{s})(p)\}$.

Example: one hundred prisoners and a lightbulb (later)

Belief revision

In dynamic epistemic logic, once you believe a fact, you will always believe it (or you will go mad: *lose* your mind).

In *belief revision*, you may *change* your mind:
first you believe p, after revision with $\neg p$ you believe $\neg p$.

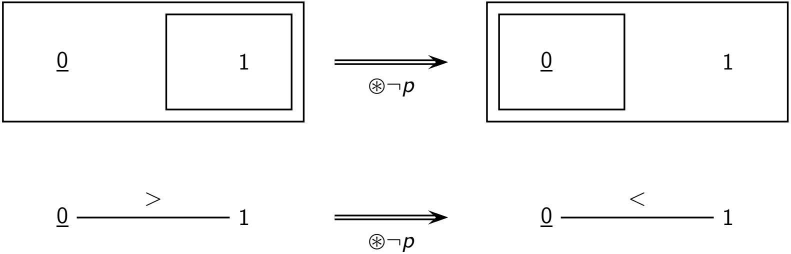
Consider Kripke/epistemic models with *preference relations*.
A *preference epistemic model* is a structure $M = \langle S, R, V, < \rangle$

- *domain* S is a nonempty set of states;
- R yields an *accessibility relation* $R_a \subseteq S \times S$ for every $a \in A$;
- *valuation* (function) $V : P \to \mathcal{P}(S)$.
- $<$ yields a *preference relation* $<_a(s) \subseteq S \times S$ for $a \in A$. For knowledge and belief, $<_a(s)$ is independent from state s, a total preorder, and we write $<_a$ (and by abbr. $\leq_a, =_a$, ...).

You believe φ, if φ is true in all *preferred* accessible states.
You know φ, if φ is true in all accessible states.

Belief revision – example

First you believe p, and after belief revision with $\neg p$ you believe $\neg p$.

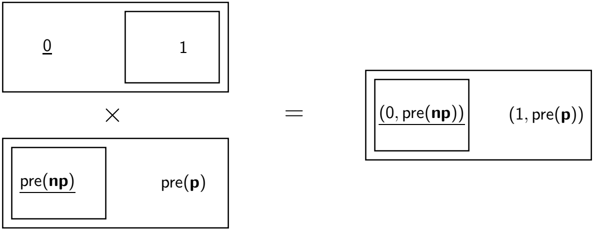


Belief revision – action models with preferences

A *preference action model* M is a structure $\langle \mathsf{S}, \sim, \mathsf{pre}, < \rangle$

- S is a *finite* domain of action points or events
- $\sim_a$ is an equivalence relation on S
- pre : S $\rightarrow \mathcal{L}$ is a precondition function that assigns a precondition to each s $\in$ S.
- $<$ yields a *preference relation* $<_a \subseteq$ S $\times$ S for $a \in A$, s $\in$ S.

Example: *soft update*. (Public announcement: *hard update*.)
First you believe p, after belief revision with $\neg p$ you believe $\neg p$.


$\underline{0}$
1
$\times$
$\underline{\mathsf{pre}(\mathbf{np})}$
$\mathsf{pre}(\mathbf{p})$
$=$
$\underline{(0, \mathsf{pre}(\mathbf{np}))}$
$(1, \mathsf{pre}(\mathbf{p}))$

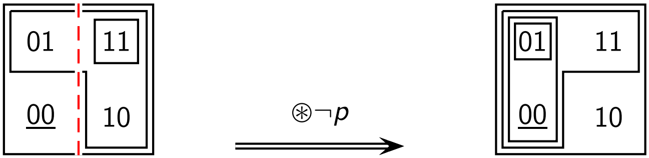
Belief revision – from update to upgrade

Given are a preference epistemic state (M, s) with $M = \langle S, \sim, V, < \rangle$ and a preference action model (M, s) with $\mathsf{M} = \langle \mathsf{S}, \sim, \mathsf{pre}, < \rangle$. The result of executing (M, s) in (M, s) is $(M \otimes \mathsf{M}, (s, \mathsf{s}))$ where $M \otimes \mathsf{M} = \langle S', \sim', V', <' \rangle$ such that:

- $S' = \{(s, \mathsf{s}) \mid s \in S, \mathsf{s} \in \mathsf{S}, \text{ and } M, s \models \mathsf{pre}(\mathsf{s})\}$
- $(s, \mathsf{s}) \sim'_a (t, \mathsf{t})$ iff $(s \sim_a t$ and $\mathsf{s} \sim_a \mathsf{t})$
- $(s, \mathsf{s}) \in V'(p)$ iff $s \in V(p)$
- $(s, \mathsf{s}) <'_a (t, \mathsf{t})$ iff $\mathsf{s} <_a \mathsf{t}$ or $(\mathsf{s} =_a \mathsf{t}$ and $s <_a t)$

Belief: true in all preferred states
Knowledge: true in all accessible states

Belief revision — example



Dynamic and temporal epistemic logic

Executing an action is like time moving on:
Dynamic epistemic logic and temporal epistemic logic are related.

A player can choose which card to show to another player:
The relation is with branching time temporal logic.
Sequences of actions correspond to histories.
Accessibility satisfies *synchronicity*, *perfect recall*, *no miracles*:

Synchronicity:
Indistinguishable sequences of actions are of equal length;
Perfect recall:
If sequences of $n + 1$ actions are indistinguishable,
the sequences of the first n actions are also indistinguishable.
No miracles:
If sequences of actions are ind. and actions are ind. then the lengthened sequences are ind.

Dynamic and temporal epistemic logic – protocols

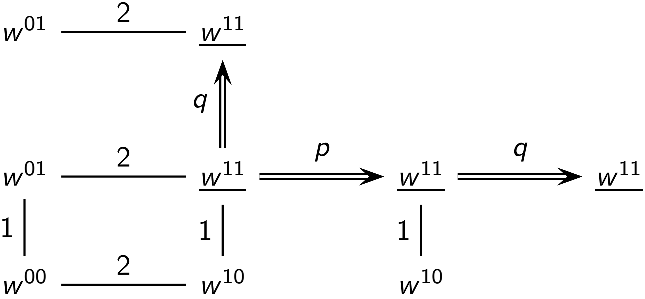
You may wish to constrain what actions are possible:
Even if you have the red card, you may not be allowed to show it;
Anne sees that Bill is muddy, but she may not announce it.
She may only announce if she knows whether she is muddy.

The allowed actions are prescribed in a *protocol*:
a prefix-closed set of sequences of actions.

Dynamic epistemic and temporal epistemic logic – forest

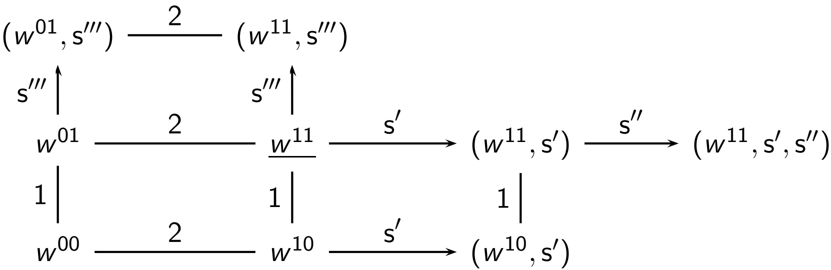
Given an epistemic model, and a protocol, we can grow a *forest*.

Example: agent 1 knows whether p, agent 2 knows whether q.
The allowed announcements are: q, p, ‘first p then q’.



Dynamic epistemic and temporal epistemic logic – forest

Forest consisting of four trees.
The protocol is $\{\mathsf{s}''', \mathsf{s}', \mathsf{s}'\mathsf{s}''\}$. (I.e.: q, p, $p; q$)



In the most basic approach, expressions like $[p][q]C_{12}(p \wedge q)$ are translated with *labelled* temporal operators, i.e., as $X_{\mathsf{s}'}X_{\mathsf{s}''}C_{12}(p \wedge q)$. There are also approaches with full-fledged future and past operators.

History

Factual change:

- van Ditmarsch, van der Hoek, Kooi 2005:
  Dynamic epistemic logic with assignment
- van Benthem, van Eijck, Kooi 2006:
  Logics of communication and change

Belief revision:

- van Ditmarsch 2005:
  Prolegomena to Dynamic Logic for Belief Revision
- Baltag & Smets 2006:
  Dynamic Belief Revision over Multi-Agent Plausibility Models

Dynamic and temporal epistemic logic:

- van Ditmarsch, van der Hoek, Ruan 2007:
  Model checking dynamic epistemics in branching time (FAMAS)
- van Benthem, Gerbrandy, Hoshi, Pacuit 2009:
  Merging frameworks for interaction

Logic puzzles

Logic puzzles and security protocols

- Russian Cards
- One hundred prisoners and a lightbulb

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.

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.

Bad:
Alice says "I have 012, or Bob has 012," and
Bob then says "I have 345, or Alice has 345."
Good:
Alice says "I have one of 012, 034, 056, 135, 246," and
Bob then says "Eve has card 6."

Card deals

Structures (interpreted system, Kripke model, state transition s.)

Players only know their own cards.
A hand of cards is a local state.
A deal of cards is a global state.

Logic (public announcement logic)

q_a		agent a holds card q.
ijk_a	$(i_a \wedge j_a \wedge k_a)$	agent a's hand of cards is $\{i, j, k\}$.

Epistemic postconditions

Bob informs Alice	aknowsbs	$\bigwedge(ijk_b \rightarrow K_a ijk_b)$
Alice informs Bob	bknowsas	$\bigwedge(ijk_a \rightarrow K_b ijk_a)$
Eve remains ignorant	cignorant	$\bigwedge(\neg K_c q_a \wedge \neg K_c q_b)$

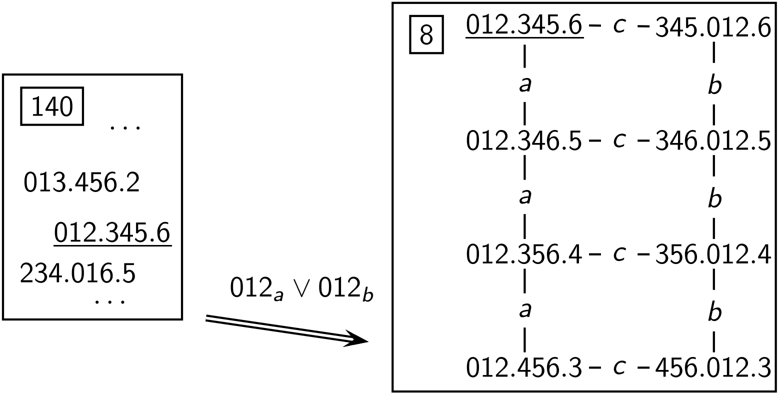
Public communication of secrets: bad

An insider says "Alice has $\{0,1,2\}$ *or Bob has* $\{0,1,2\}$*."*

$$012.345.6 \models [012_a \vee 012_b]\text{cignorant}$$

Alice says "I have $\{0,1,2\}$ *or Bob has* $\{0,1,2\}$*."*

$$012.345.6 \not\models [K_a(012_a \vee 012_b)]\text{cignorant}$$



140
. . .
013.456.2
012.345.6
234.016.5
. . .
$012_a \vee 012_b$
8
012.345.6 – c – 345.012.6
a
b
012.346.5 – c – 346.012.5
a
b
012.356.4 – c – 356.012.4
a
b
012.456.3 – c – 456.012.3

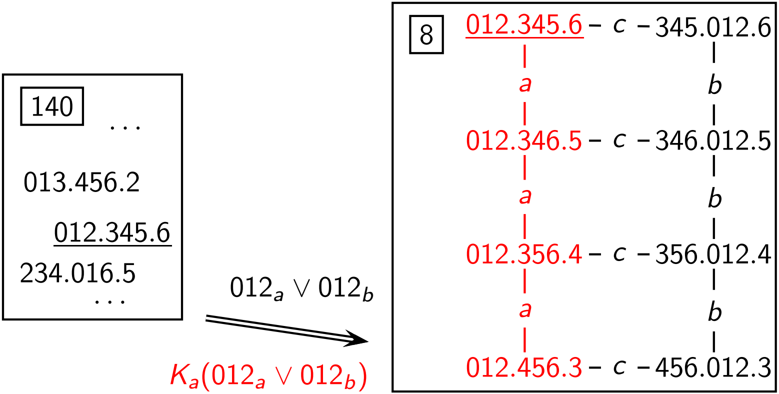
Public communication of secrets: bad

An insider says "Alice has $\{0,1,2\}$ *or Bob has* $\{0,1,2\}$*."*

$$012.345.6 \models [012_a \vee 012_b]\text{cignorant}$$

Alice says "I have $\{0,1,2\}$ *or Bob has* $\{0,1,2\}$*."*

$$012.345.6 \not\models [K_a(012_a \vee 012_b)]\text{cignorant}$$



$\boxed{140}$

. . .

013.456.2

012.345.6

234.016.5

. . .

$012_a \vee 012_b$

$K_a(012_a \vee 012_b)$

$\boxed{8}$

012.345.6 – c – 345.012.6

a b

012.346.5 – c – 346.012.5

a b

012.356.4 – c – 356.012.4

a b

012.456.3 – c – 456.012.3

Public communication of secrets: also bad

Alice says "I don't have card 6."

$$\begin{array}{l} 012.345.6 \models [K_a \neg 6_a]\mathrm{cignorant} \\ 012.345.6 \not\models [K_a \neg 6_a] K_a \mathrm{cignorant} \end{array}$$

Public communication of secrets: almost good

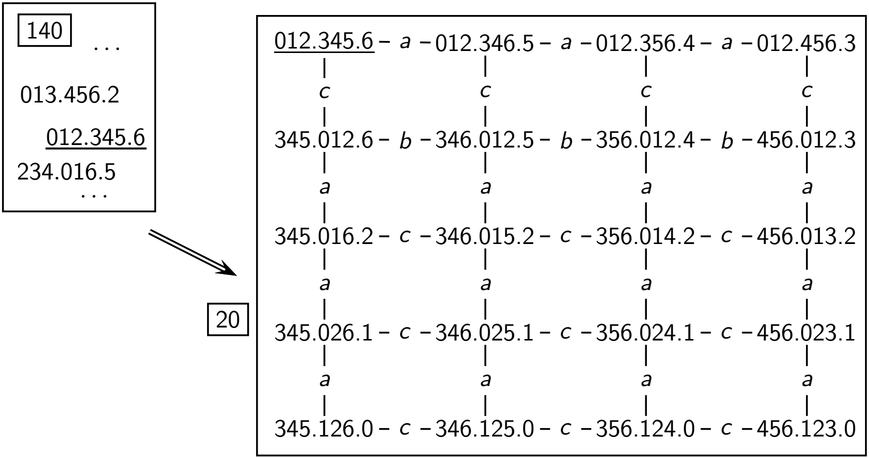
Alice says "I have $\{0,1,2\}$*, or I have none of these cards."*
Eve is ignorant after Alice's announcement.
Alice knows that Eve is ignorant.
Eve doesn't know that Alice knows that Eve is ignorant.
But Eve may assume that Alice knows that Eve is ignorant.
That is informative for Eve!

$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]\text{cignorant}$$
$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]K_a\text{cignorant}$$
$$012.345.6 \not\models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]K_c K_a\text{cignorant}$$
$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))][K_a\text{cignorant}]\neg\text{cignorant}$$

$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))][K_a\text{cignorant}]\neg K_a\text{cignorant}$$

Alice reveals her cards, *because* she intends to keep them secret.

Public communication of secrets: almost good



140

. . .

013.456.2

012.345.6

234.016.5

. . .

20

012.345.6 – a – 012.346.5 – a – 012.356.4 – a – 012.456.3

c c c c

345.012.6 – b – 346.012.5 – b – 356.012.4 – b – 456.012.3

a a a a

345.016.2 – c – 346.015.2 – c – 356.014.2 – c – 456.013.2

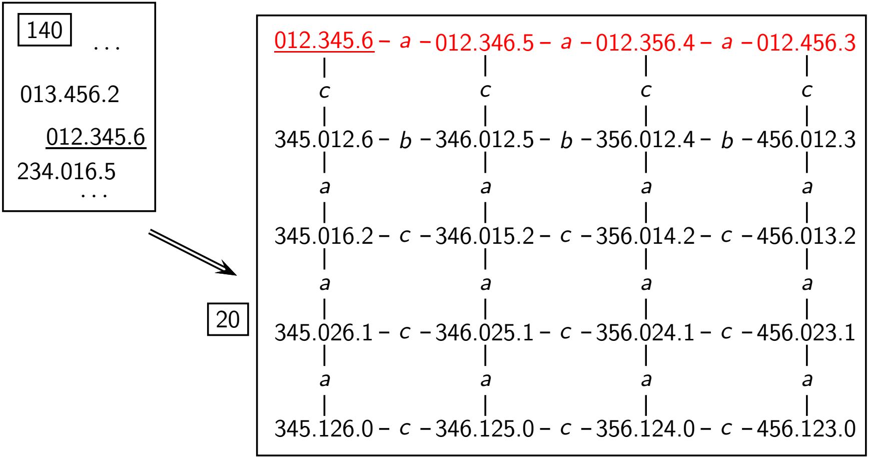
a a a a

345.026.1 – c – 346.025.1 – c – 356.024.1 – c – 456.023.1

a a a a

345.126.0 – c – 346.125.0 – c – 356.124.0 – c – 456.123.0

Public communication of secrets: almost good



140

. . .

013.456.2

<u>012.345.6</u>

234.016.5

. . .

20

<u>012.345.6</u> – *a* – 012.346.5 – *a* – 012.356.4 – *a* – 012.456.3

c *c* *c* *c*

345.012.6 – *b* – 346.012.5 – *b* – 356.012.4 – *b* – 456.012.3

a *a* *a* *a*

345.016.2 – *c* – 346.015.2 – *c* – 356.014.2 – *c* – 456.013.2

a *a* *a* *a*

345.026.1 – *c* – 346.025.1 – *c* – 356.024.1 – *c* – 456.023.1

a *a* *a* *a*

345.126.0 – *c* – 346.125.0 – *c* – 356.124.0 – *c* – 456.123.0

Public communication of secrets

Safe announcements guarantee public preservation of ignorance.

$[\varphi]$	announcement of φ (by an observer)
$[K_a\varphi]$	announcement of φ (by agent/Alice)
$[K_a\varphi \wedge [K_a\varphi]C_{abc}\text{cignorant}]$	safe announcement of φ
$[K_a\varphi][C_{abc}\text{cignorant}]$	

Good protocols produce finite sequences of safe announcements s.t.

$$C_{abc}(\text{aknowsbs} \wedge \text{bknowsas} \wedge \text{cignorant})$$

Now for some good solutions!

One hundred prisoners and a lightbulb

A group of 100 prisoners, all together in the prison dining area, are told that they will be all put in isolation cells and then will be interrogated one by one in a room containing a light with an on/off switch. The prisoners may communicate with one another by toggling the light-switch (and that is the only way in which they can communicate). The light is initially switched off. There is no fixed order of interrogation, or interval between interrogations, and the same prisoner may be interrogated again at any stage. When interrogated, a prisoner can either do nothing, or toggle the light-switch, or announce that all prisoners have been interrogated. If that announcement is true, the prisoners will (all) be set free, but if it is false, they will all be executed. While still in the dining room, and before the prisoners go to their isolation cells (forever), can the prisoners agree on a protocol that will set them free?

100 prisoners — solution

Protocol for $n > 3$ prisoners:

The n prisoners appoint one amongst them as the counter. All non-counting prisoners follow the following protocol: the first time they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first $n - 2$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

100 prisoners — solution

Protocol for $n > 3$ prisoners:

The n prisoners appoint one amongst them as the counter. All non-counting prisoners follow the following protocol: the first time they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first $n - 2$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

What if it is not known whether the light is initially on?

100 prisoners — solution

Protocol for $n > 3$ prisoners:

The n prisoners appoint one amongst them as the counter. All non-counting prisoners follow the following protocol: the first time they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first $n - 2$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

What if it is not known whether the light is initially on?
Same count, you may get hanged (namely if light was on).
One higher, you may never terminate (namely if light was off).
???

100 prisoners — solution if light may be on or off

The n prisoners appoint one amongst them as the counter. All non-counting prisoners follow the following protocol: the ~~first time~~ first two times they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first ~~$n-2$~~ $2n-3$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

100 prisoners — solution if light may be on or off

The n prisoners appoint one amongst them as the counter. All non-counting prisoners follow the following protocol: the ~~first time~~ first two times they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first ~~$n-2$~~ $2n-3$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.

For $n = 100$, the next entry (198) after 197 switches:

light was off and 99 non-counters have been interrogated twice
light was on and 98 non-counters twice and one once only.

Either way is fine!

100 prisoners — knowing before the counter

After a non-counter has turned the light on, he counts the number of times he sees the sequence ‘light off – light on’.

If this is 98 times, all have been interrogated.

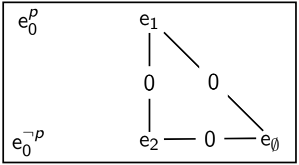
His announcement will then be before the counter’s.

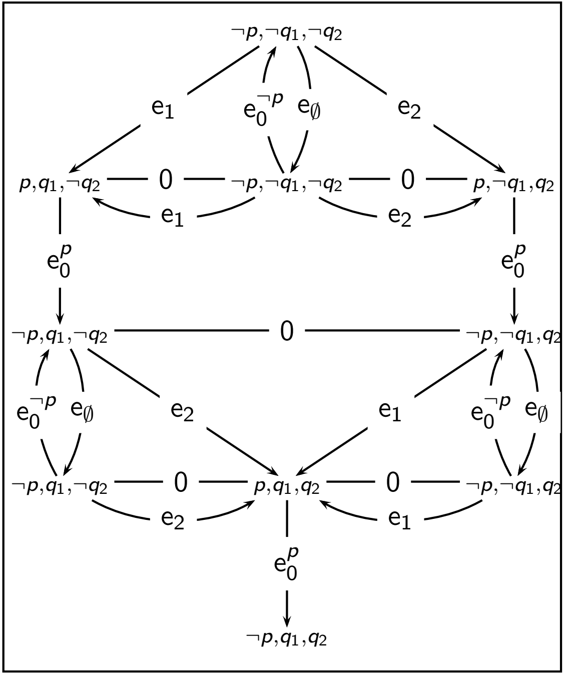
3 prisoners — a dash of logic

Proposition p stands for for ‘the light is on’.
The counter is agent 0. The non-counters are not modelled.
Proposition q_1 stands for ‘prisoner 1 has turned on the light’.
Proposition q_2 stands for ‘prisoner 2 has turned on the light’.

event	*precond.*	*postcondition*
$\mathsf{e}_\emptyset$	if $\top$	then ϵ ‘nothing happens’
e_1	if $\top$	then $p := q_1 \rightarrow p$ and $q_1 := p \rightarrow q_1$
e_2	if $\top$	then $p := q_2 \rightarrow p$ and $q_2 := p \rightarrow q_2$
$\mathsf{e}_0^{\neg p}$	if $\neg p$	then ϵ
e_0^{p}	if p	then $p := \bot$

How the events appear to agent 0:




$\neg p, \neg q_1, \neg q_2$
e_1
$e_0^{\neg p}$
$e_\emptyset$
e_2
$p, q_1, \neg q_2$
0
$\neg p, \neg q_1, \neg q_2$
0
$p, \neg q_1, q_2$
e_1
e_2
e_0^p
e_0^p
$\neg p, q_1, \neg q_2$
0
$\neg p, \neg q_1, q_2$
$e_0^{\neg p}$
$e_\emptyset$
e_2
e_1
$e_0^{\neg p}$
$e_\emptyset$
$\neg p, q_1, \neg q_2$
0
p, q_1, q_2
0
$\neg p, \neg q_1, q_2$
e_2
e_1
e_0^p
$\neg p, q_1, q_2$

100 prisoners — synchronization

Assume a single interrogation per day takes place.
When can the prisoners expect to be set free from prison?

100 prisoners — synchronization

Assume a single interrogation per day takes place.
When can the prisoners expect to be set free from prison?

non-counter / counter / another non-counter / counter / etc.

$\frac{99}{100}$ / $\frac{1}{100}$ / $\frac{98}{100}$ / $\frac{1}{100}$ / etc.

$\frac{100}{99}$ / $\frac{100}{1}$ / $\frac{100}{98}$ / $\frac{100}{1}$ / etc.

100 prisoners — synchronization

Assume a single interrogation per day takes place.
When can the prisoners expect to be set free from prison?

non-counter / counter / another non-counter / counter / etc.

$\frac{99}{100}$ / $\frac{1}{100}$ / $\frac{98}{100}$ / $\frac{1}{100}$ / etc.

$\frac{100}{99}$ / $\frac{100}{1}$ / $\frac{100}{98}$ / $\frac{100}{1}$ / etc.

Summation:

$$\sum_{i=1}^{99}\left(\frac{100}{i}+\frac{100}{1}\right) = 99\cdot 100 + 100\cdot \sum_{i=1}^{99} \frac{1}{i} = 9,900 + 518 \text{ days } \approx 28.5 \text{ years}$$

100 prisoners — improvements given synchronization

Dynamic counter assignment (protocol in two stages):

- stage 1, 99 days: the first prisoner to enter the room twice turns on the light. (Expectation: 13 days.)
- stage 1, day 100: if light off, done; otherwise, turn light off.
- stage 2, from day 101: as before, except that:
  counter twice interrogated on day n counts until $100 - n$ only;
  non-counters who only saw light off in stage 1: do nothing;
  non-counters who saw light on in stage 1: do the usual. (24 y)

Head counter and assistant counters (iterated protocol, 2 stages):

- stage 1: head and assistant counters count to agreed max. n;
- stage 2: head counter collects from *successful* assistants;
- repeat stage 1 (unsuccessful assistants continue counting to n) and stage 2 (not yet collected successful assistants, and newly successful assistants) until termination. (9 years)

Minimum not known!

Quantifying over information change

Modal logics with implicit propositional quantification.