ESSLLI08 Hamburg: Dynamic Epistemic Logic

lecturers: Hans van Ditmarsch & Jan van Eijck

Hans' part of the course: mornings & Friday

- Monday morning: epistemic logic
- Tuesday morning: public announcements
- Wednesday morning: action models
- Thursday morning: factual change
- Friday: logic puzzles & security

`hans@cs.otago.ac.nz`
`http://www.cs.otago.ac.nz/staffpriv/hans/`

Epistemic Logic

Ia: 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: $K_a 0_a$
- 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: $\hat{K}_a \neg 1_t$
- 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: $K_a 0_a \vee K_a 1_a \vee K_a 2_a$

Structures

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

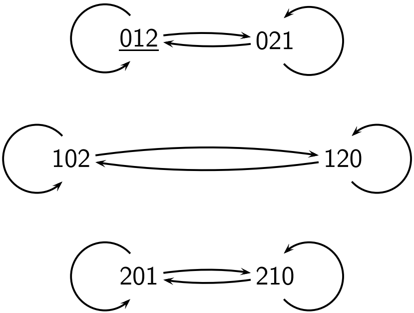
- *domain* S is a nonempty set of states;
- R yields an *accessibility relation* $R_a \subseteq S \times S$ for every $a \in A$;
- *valuation* (function) $V : P \to \mathcal{P}(S)$.

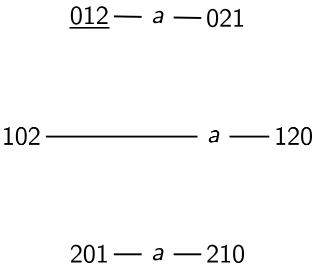
If all the relations R_a in M are equivalence relations, we call M an *epistemic model*. In that case, we write $\sim_a$ rather than R_a, and we represent the model as $M = \langle S, \sim, V \rangle$.

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

Example

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





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

$Hexa1, 012 \models K_a 0_a$

$\Leftrightarrow$

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

$\Leftarrow$

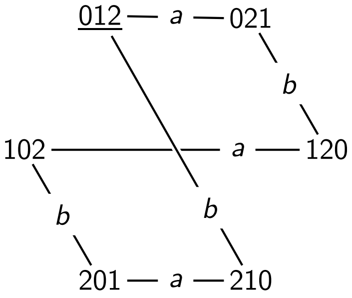
$Hexa1, 012 \models 0_a$ and $Hexa1, 021 \models 0_a$

$\Leftrightarrow$

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

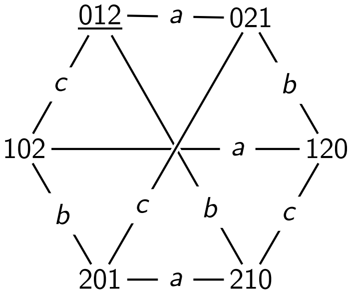
Two agents

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



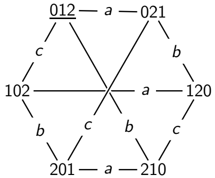
- Bill does not consider it possible that Anne has card 1: $\neg \hat{K}_b 1_a$
- Anne considers it possible that Bill considers it possible that she has card 1: $\hat{K}_a \hat{K}_b 1_a$
- Anne knows Bill to consider it possible that she has card 0: $K_a \hat{K}_b 0_a$

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



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

$\Leftarrow$ because $012 \sim_a 021$

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

$\Leftarrow$ because $021 \sim_b 120$

$Hexa, 120 \models K_c \neg 0_a$

$\Leftrightarrow$ $\sim_c(120) = \{120, 210\}$

$Hexa, 120 \models \neg 0_a$ and $Hexa, 210 \models \neg 0_a$

$\Leftrightarrow$

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

$\Leftrightarrow$

$120, 210 \notin V_{0_a} = \{012, 021\}$

Properties of knowledge

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

Realistic assumptions for knowledge?

Axiomatization

all instantiations of propositional tautologies
$K_a(\varphi \rightarrow \psi) \rightarrow (K_a\varphi \rightarrow K_a\psi)$
$K_a\varphi \rightarrow \varphi$
$K_a\varphi \rightarrow K_aK_a\varphi$
$\neg K_a\varphi \rightarrow K_a\neg K_a\varphi$
From φ and $\varphi \rightarrow \psi$, infer ψ
From φ, infer $K_a\varphi$

History

- von Wright 1951: An Essay in Modal Logic
- Hintikka 1962: Knowledge and Belief
- Aumann 1976: Agreeing to Disagree
- Fagin, Halpern, Moses and Vardi 1995: Reasoning about Knowledge
- Meyer and van der Hoek 1995: Epistemic Logic for AI and Computer Science

Common knowledge

Ib: Common knowledge

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:
$E_G\varphi := K_1\varphi \wedge K_2\varphi \wedge \ldots \wedge K_{\text{last}}\varphi$

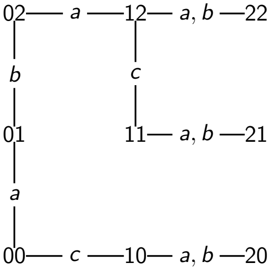
Common knowledge:
$C_G\varphi := \varphi \wedge E_G\varphi \wedge E_G E_G\varphi \wedge \ldots$
or
$C_G\varphi := \varphi \wedge K_1\varphi \wedge K_2\varphi \wedge K_1 K_1\varphi \wedge K_1 K_2\varphi \wedge \ldots K_1 K_1 K_1\varphi \ldots$

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

Computing transitive closure

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

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.



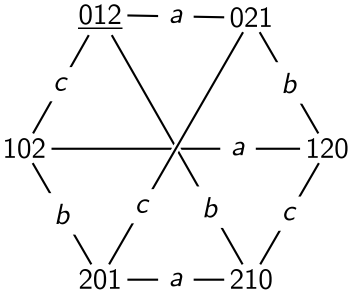
What is the partition on these nine states for a?
For group $\{a, b\}$? For group $\{a, c\}$? For group $\{a, b, c\}$?

Epistemic Logic with Common Knowledge

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

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

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

$$\begin{tikzcd}
02 \arrow[r, dash, "a"] \arrow[d, dash, "b"'] & 12 \arrow[r, dash, "{a, b}"] \arrow[d, dash, "c"] & 22 \\
01 \arrow[d, dash, "a"'] & 11 \arrow[r, dash, "{a, b}"] & 21 \\
00 \arrow[r, dash, "c"] & 10 \arrow[r, dash, "{a, b}"] & 20
\end{tikzcd}$$

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 $C_B\varphi$

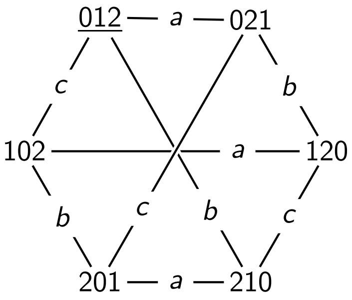
History

- 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

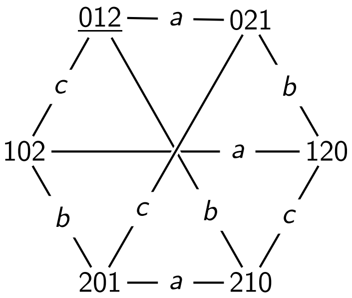
II: Public announcements

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 Announcements: language

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

Public Announcements: 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 $[\varphi]$.

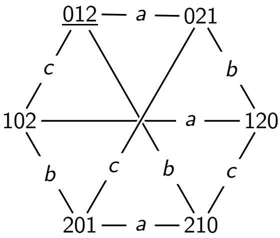
'φ is the announcement' means 'φ is publicly and truthfully announced'.

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

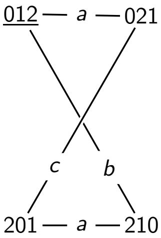
$M|\varphi := \langle S', \sim', V' \rangle$:

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

Example announcement in Hexa



$\Rightarrow$



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

$$\Leftrightarrow$$

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

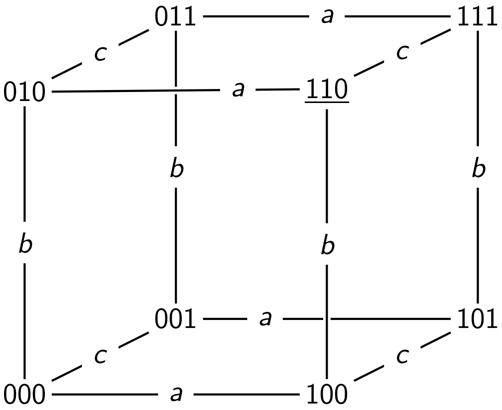
$$\Leftarrow \qquad\qquad \sim_c(012) = \{012\}$$

$$012 \neq V(1_a) \text{ and } Hexa|\neg 1_a, 012 \models 0_a$$

Muddy Children

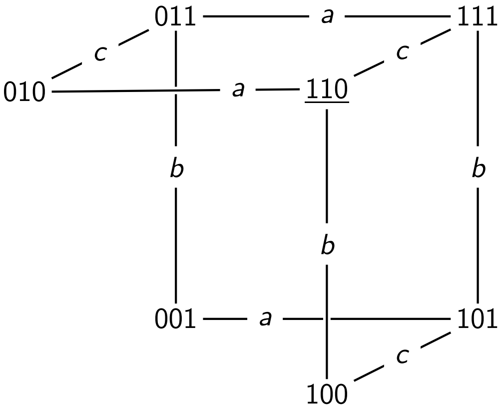
A group of children has been playing outside and are called back into the house by their father. The children gather round him. As one may imagine, some of them have become dirty from the play and in particular: they may have mud on their forehead. Children can only see whether other children are muddy, 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



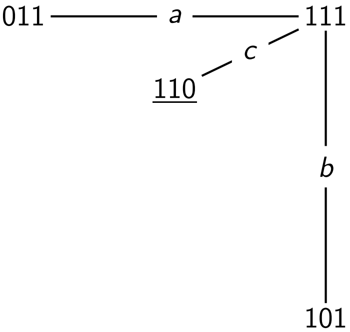
Given: The children can see each other

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

<u>110</u>

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

Theorem (Plaza, Gerbrandy)

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

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

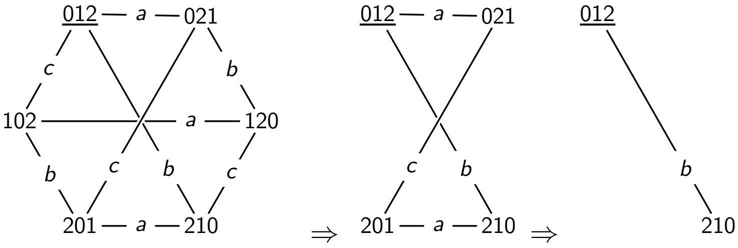
Sequence of announcements

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:

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



Unsuccessful updates

Postulate of success:

$$\varphi \rightarrow \langle \varphi \rangle C_A \varphi$$

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:

$$\models [p \wedge \neg K_b p] \neg (p \wedge \neg K_b p)$$

$$0 \overset{a}{\text{———————}} \underline{1} \qquad \xRightarrow[p \wedge \neg K_a p]{} \qquad \underline{1}$$

History

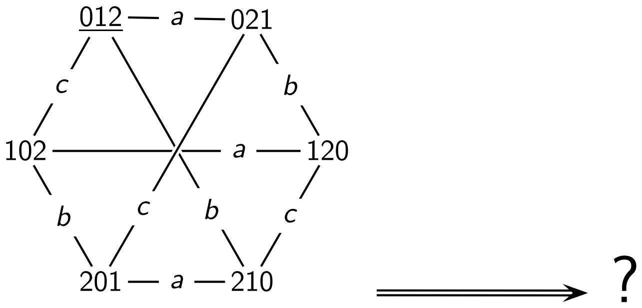
- Plaza 1989: Logics of Public Communications
- Gerbrandy & Groeneveld 1997: Reasoning about Information Change
- Baltag, Moss & Solecki 1998: The Logic of Common Knowledge, Public Announcements, and Private Suspicions
- van Ditmarsch, van der Hoek & Kooi 2007: Dynamic Epistemic Logic

Action models

III: Action models

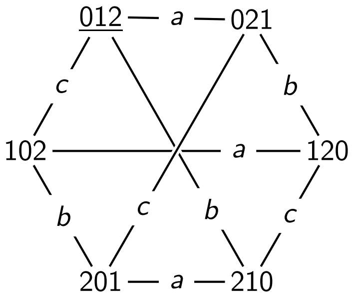
What we cannot do yet...

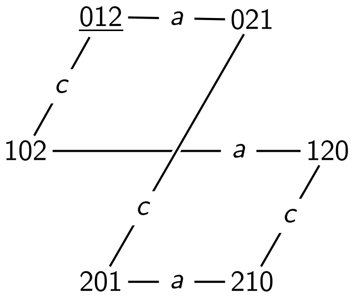
(Anne holds 0, Bill holds 1, and Cath holds 2.) Anne shows (only) Bill card 0. Cath cannot see the face of the shown card, but notices that a card is being shown.



What we cannot do yet...

(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.





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

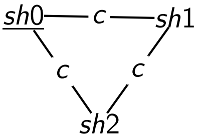
Action models

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
- pre : $\mathsf{S} \rightarrow \mathcal{L}$ is a preconditions function that assigns a precondition to each $\mathsf{s} \in \mathsf{S}$.

Showing a card

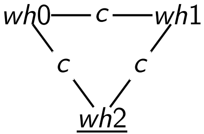
(Anne holds 0, Bill holds 1, and Cath holds 2.) Anne shows (only) Bill her card. (It is card 0.) Cath cannot see the face of the shown card, but notices that a card is being shown.



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

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.



- $\mathsf{S} = \{\mathsf{wh0}, \mathsf{wh1}, \mathsf{wh2}\}$
- $\sim_a = \{(\mathsf{s}, \mathsf{s}) \mid \mathsf{s} \in \mathsf{S}\}$, $\sim_b = \{(\mathsf{s}, \mathsf{s}) \mid \mathsf{s} \in \mathsf{S}\}$, $\sim_c = \mathsf{S} \times \mathsf{S}$
- $\mathsf{pre}(\mathsf{sh0}) = \neg 0_a$, $\mathsf{pre}(\mathsf{sh1}) = \neg 1_a$, $\mathsf{pre}(\mathsf{sh2}) = \neg 2_a$

What do you learn from an action?

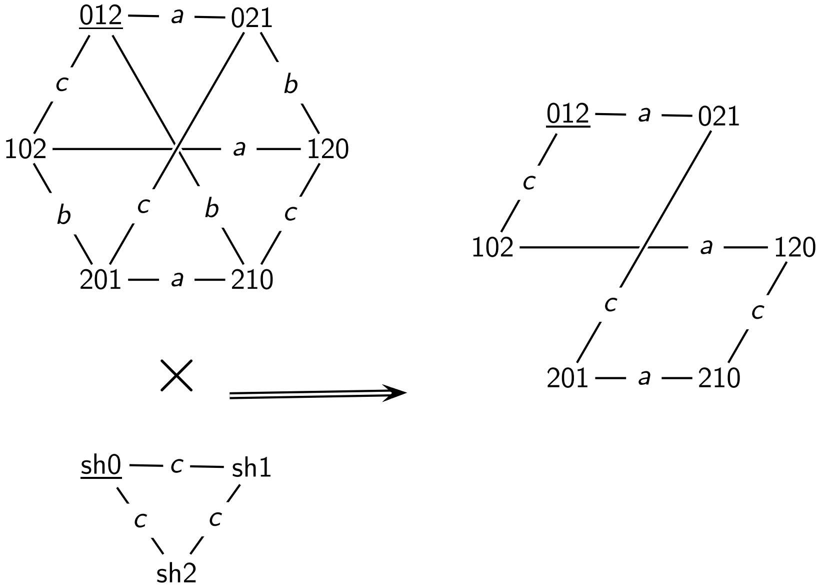
- Firstly, if you can distinguish two actions, then you can also distinguish the states that result from executing the action.
- Secondly, you do not forget anything due to an action. States that you could distinguish before an action are still distinguishable.

Product update

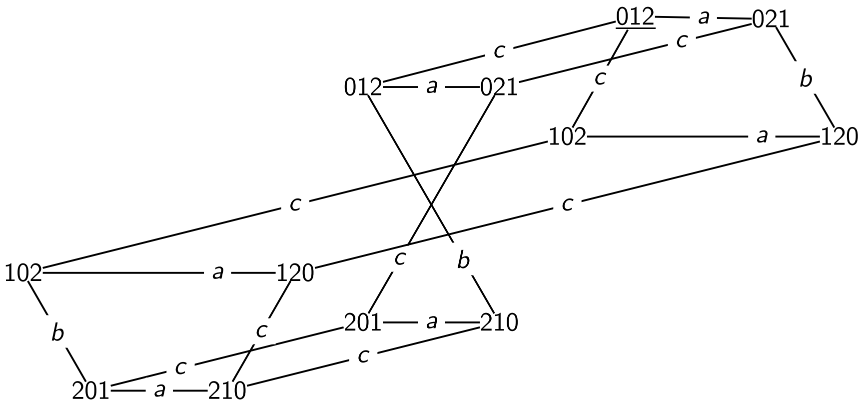
Given are an epistemic state (M, s) with $M = \langle S, \sim, V\rangle$ and an action model (M, s) with $\mathsf{M} = \langle \mathsf{S}, \sim, \mathsf{pre}\rangle$. The result of executing (M, s) in (M, s) is $(M \otimes \mathsf{M}, (s, \mathsf{s}))$ where $M \otimes \mathsf{M} = \langle S', \sim', V'\rangle$ 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 $(s \sim_a t$ 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 2’ 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

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

Syntax and semantics

- Are syntax and semantics clearly separated?

YES

Axiomatization

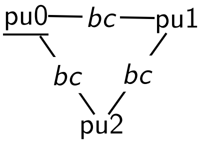
$[\mathsf{M},\mathsf{s}]p \leftrightarrow (\mathsf{pre}(\mathsf{s}) \rightarrow p)$
$[\mathsf{M},\mathsf{s}]\neg\varphi \leftrightarrow (\mathsf{pre}(\mathsf{s}) \rightarrow \neg[\mathsf{M},\mathsf{s}]\varphi)$
$[\mathsf{M},\mathsf{s}](\varphi \wedge \psi) \leftrightarrow ([\mathsf{M},\mathsf{s}]\varphi \wedge [\mathsf{M},\mathsf{s}]\psi)$
$[\mathsf{M},\mathsf{s}]K_a\varphi \leftrightarrow (\mathsf{pre}(\mathsf{s}) \rightarrow \bigwedge_{\mathsf{s}\sim_a\mathsf{t}} K_a[\mathsf{M},\mathsf{t}]\varphi)$
$[\mathsf{M},\mathsf{s}][\mathsf{M}',\mathsf{s}']\varphi \leftrightarrow [(\mathsf{M},\mathsf{s});(\mathsf{M}',\mathsf{s}')]\varphi$
From φ, infer $[\mathsf{M},\mathsf{s}]\varphi$
Let (M,s) be an action model and let a set of formulas χ_t for every t such that $\mathsf{s} \sim_B \mathsf{t}$ be given. From $\chi_\mathsf{t} \rightarrow [\mathsf{M},\mathsf{t}]\varphi$ and $(\chi_\mathsf{t} \wedge \mathsf{pre}(\mathsf{t})) \rightarrow K_a\chi_\mathsf{u}$ for every $\mathsf{t} \in \mathsf{S}$ such that $\mathsf{s} \sim_B \mathsf{t}$, $a \in B$ and $\mathsf{t} \sim_a \mathsf{u}$, infer $\chi_\mathsf{s} \rightarrow [\mathsf{M},\mathsf{s}]C_B\varphi$.

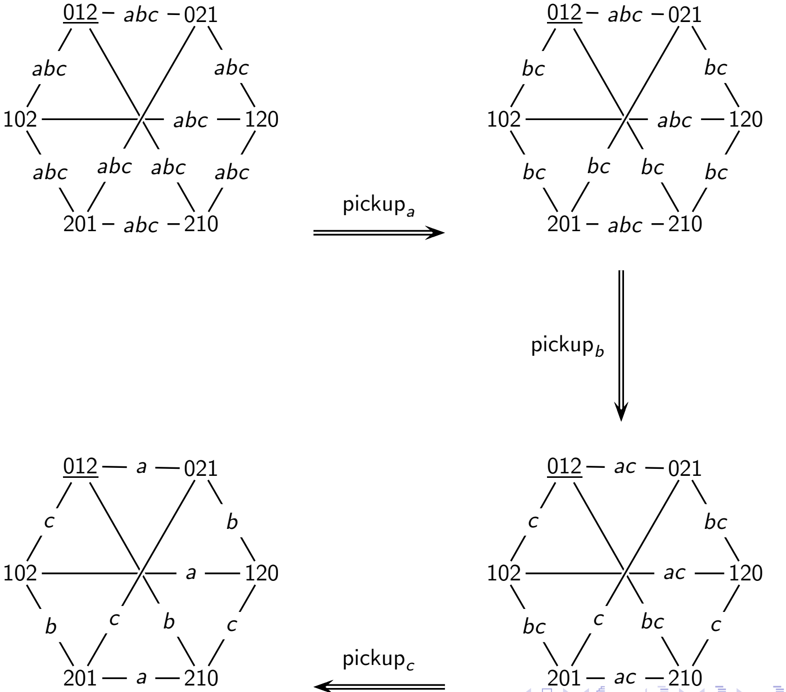
Every formula in the language of action model logic without common knowledge is equivalent to a formula in the language of epistemic logic.

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 — abc — 021
abc
abc
102 — abc — 120
abc
abc
abc
abc
201 — abc — 210
pickup_a
012 — abc — 021
bc
bc
102 — abc — 120
bc
bc
bc
bc
201 — abc — 210
pickup_b
012 — ac — 021
c
bc
102 — ac — 120
bc
c
bc
c
201 — ac — 210
pickup_c
012 — a — 021
c
b
102 — a — 120
b
c
b
c
201 — a — 210

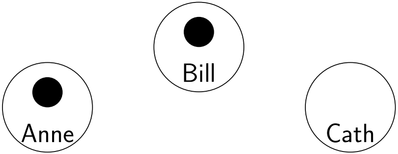
History

- Baltag, Moss & Solecki 1998: The Logic of Common Knowledge, Public Announcements, and Private Suspicions

Factual change

IV: Factual change

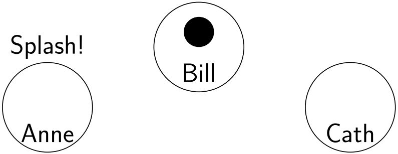
Factual change — Muddy Children again



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.)

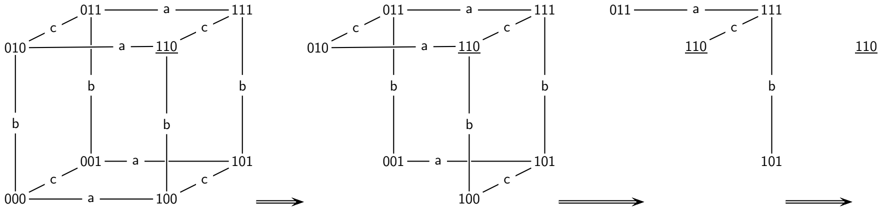
Cleaning Muddy Children



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

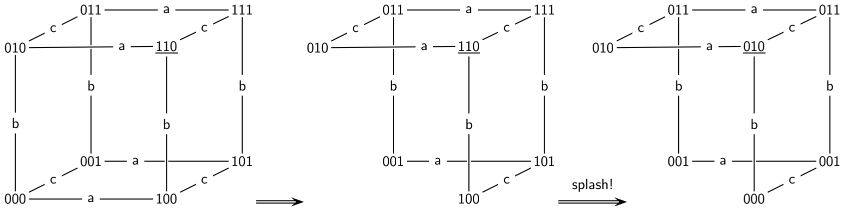
- At least one of you is muddy.
- **Splash!** *Father empties a bucket of water over Anne.*
- If you know whether you are muddy, step forward. (...?)
- If you know whether you are muddy, step forward. (...?)

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

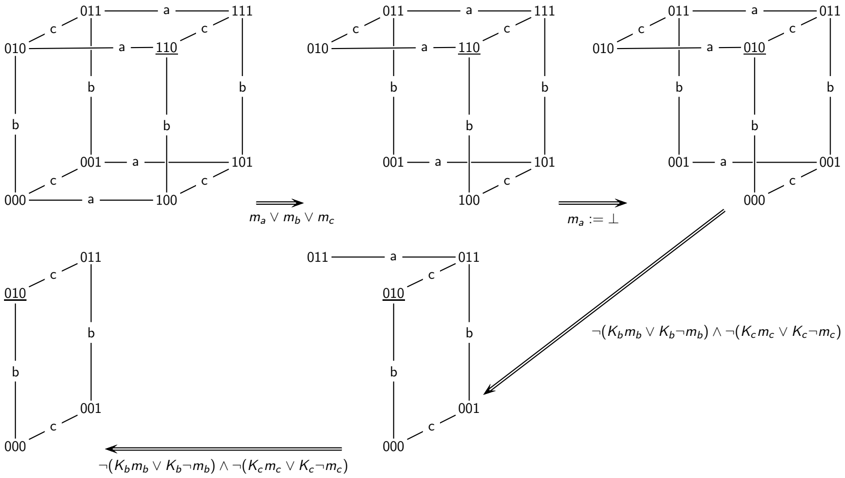
$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_a\varphi \mid C_A\varphi \mid [\varphi]\psi \mid [p := \varphi]\psi$$

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

Factual change with action models, more technique, and history: Jan

Logic puzzles

V: 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	$(i_a \land j_a \land k_a)$	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	$\bigwedge(\neg K_c q_a \land \neg K_c q_b)$

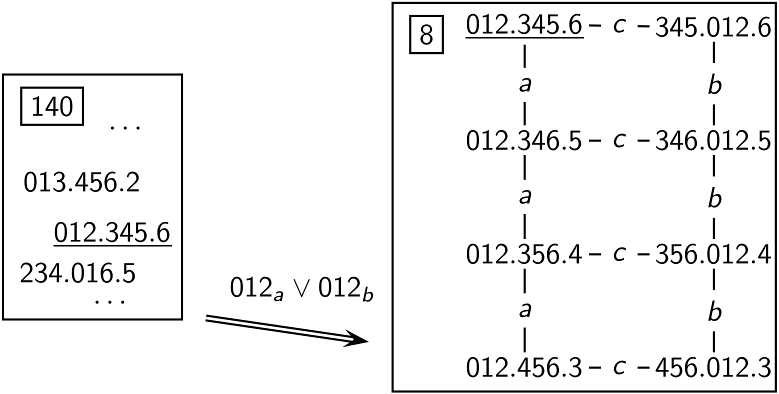
Public communication of secrets: bad

An observer 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* $\{0,1,2\}$*."*

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



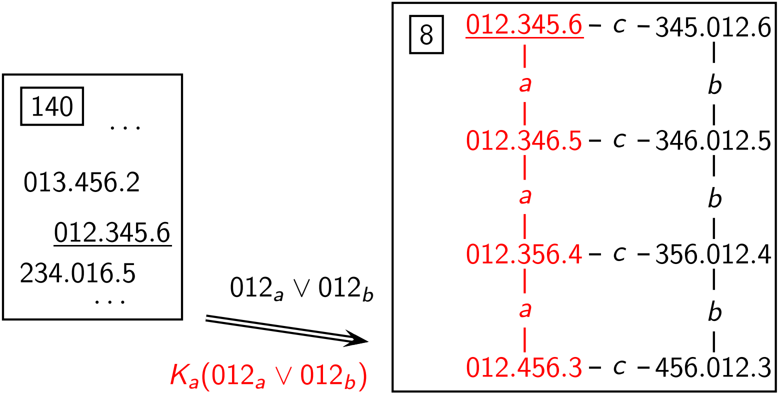
Public communication of secrets: bad

An observer 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* $\{0,1,2\}$*."*

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



Public communication of secrets: also bad

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

$$012.345.6 \models [K_a \neg 6_a] \mathrm{cignorant}$$
$$012.345.6 \not\models [K_a \neg 6_a] K_a \mathrm{cignorant}$$

Public communication of secrets: almost good

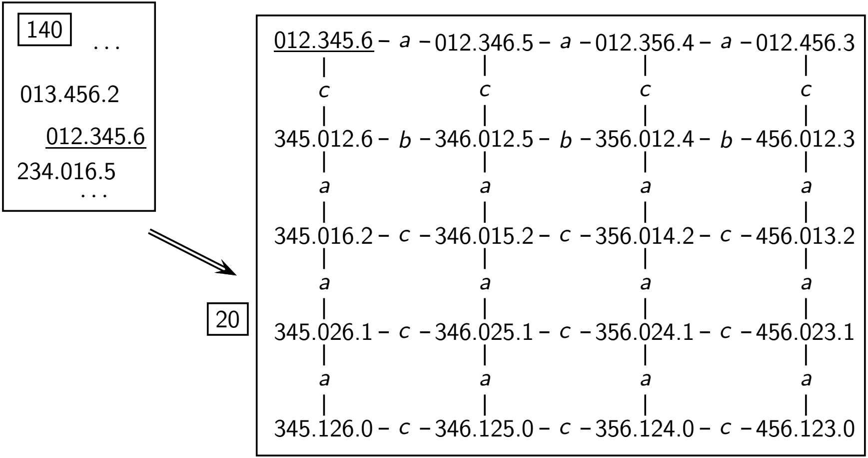
Alice says "I have $\{0, 1, 2\}$*, or I have none of these cards."*
Eve is ignorant after Alice's announcement.
Alice knows that Eve is ignorant.
Eve doesn't know that Alice knows that Eve is ignorant.
But Eve may assume that Alice knows that Eve is ignorant.
That is informative for Eve!

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

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

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

Public communication of secrets: almost good



Public communication of secrets: almost good

$\boxed{140}$ . . .

013.456.2

<u>012.345.6</u>

234.016.5

. . .

$\boxed{20}$

<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.

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

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 (assuming that at any stage every prisoner will be interrogated again sometime)?