Institute for Logic, Language and Computation

Tutorial Dynamic Logic of Information

Epistemic and Dynamic Logic

Johan van Benthem

April 2022

Tsinghua Workshop on Dynamics in Logic and Language

INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

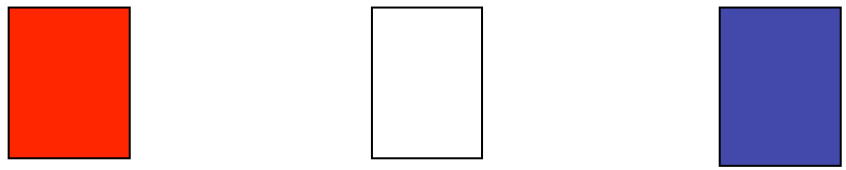
Information Flow Through

Observation and Communication

Institute for Logic, Language and Computation

Three Cards

John, Mary, Paul get one card each



John **Red** Mary **White** Paul **Blue**

Mary asks John: **Do you have the blue card?**

Who knows the deal of the cards now?

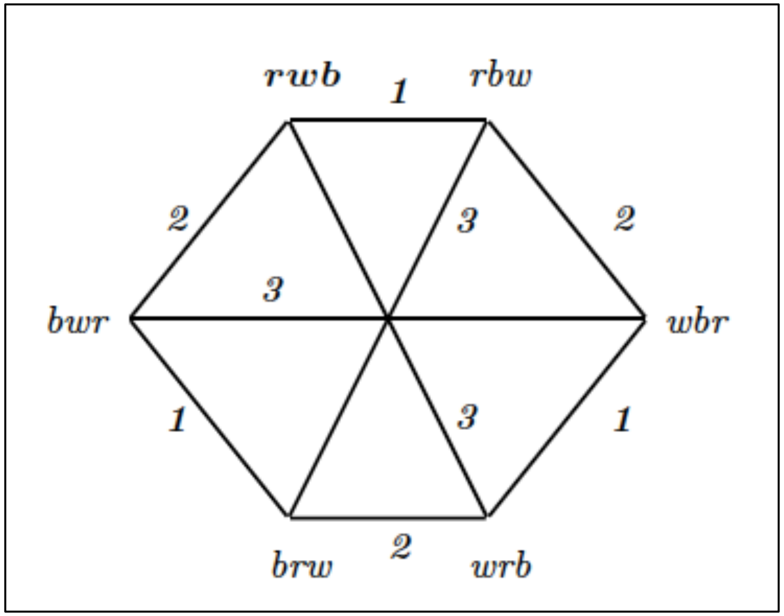
John answers: **No**.

Who knows what now?

Institute for Logic, Language and Computation

Information Flow: What do Agents Know and How Does this Knowledge Change?

Information Diagram



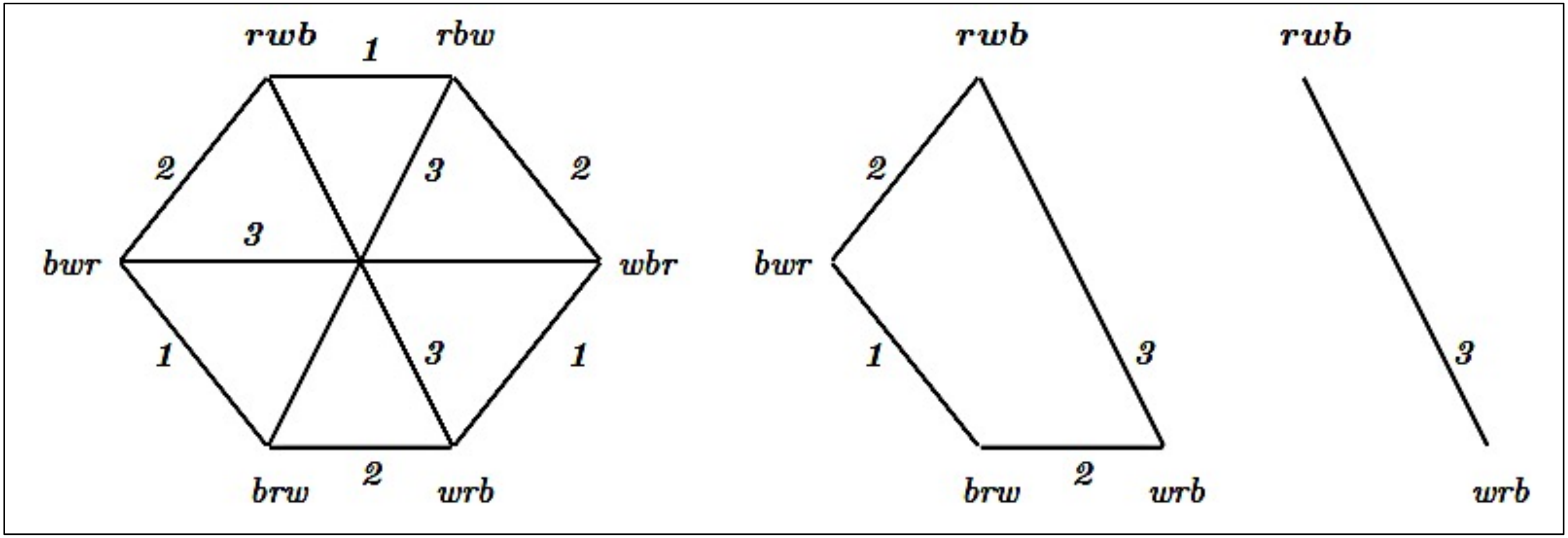
points ~ possible deals of cards

uncertainty lines ~ what agents can(not) observe

knowledge: what is true in all your current options

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Information Change and Updates



Final state John and Mary know the cards, Paul does not.

But Paul knows that the others know, and in fact,

this is common knowledge in the group

Institute for Logic, Language and Computation

The Methodology of Logic:

Formal Language and Semantics

\

Institute for Logic, Language and Computation

Knowledge, Formal and Natural Language

Language $p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_i\varphi$

existential dual modality $<i>\varphi := \neg K_i \neg\varphi$

Question: "φ?" Answer: "Yes".

$\neg K_Q\varphi$	**Q** does not know that φ
$\neg K_Q\varphi \wedge \neg K_Q\neg\varphi$	**Q** does not know whether φ
$K_Q(K_A\varphi \vee K_A\neg\varphi)$	**Q** knows that **A** knows whether φ
$<Q>(K_A\varphi \vee K_A\neg\varphi)$	…

effect afterwards: common knowledge $C_{\{Q, A\}}\varphi$

Institute for Logic, Language and Computation

Epistemic Models and Semantics

semantic information: range of options for the real world

epistemic model M = (W, $\{\sim_i\}_i$, V)

W **worlds/points**, epistemic **accessibility relations** $\sim_i$,

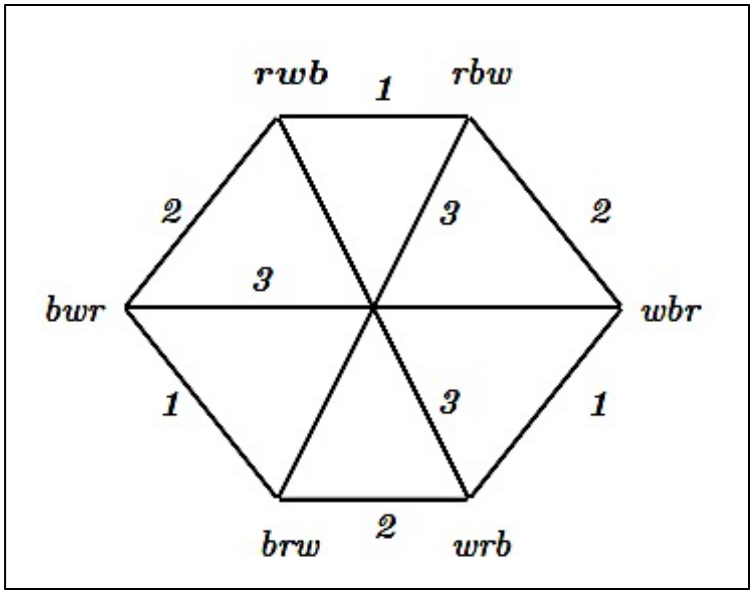
(for now: equivalence relation: reflexive, symmetric, transitive)

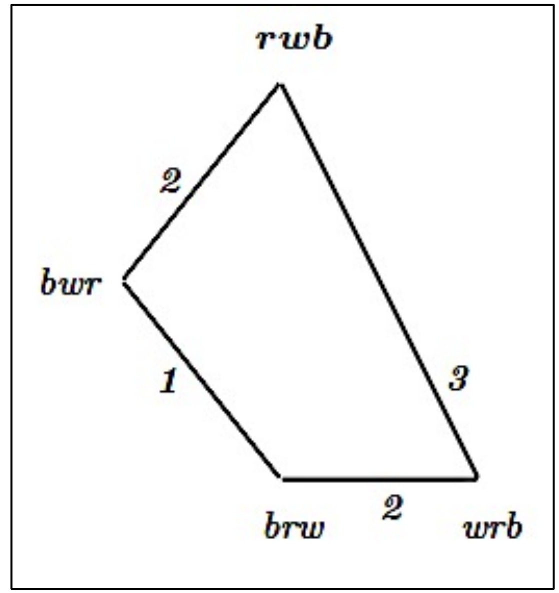
valuation V: truth values for proposition letters at worlds

truth definition **M**, s |= $K_i\varphi$ iff for all t $\sim_i$ s: **M**, t |= φ

Institute for Logic, Language and Computation

Model Checking Formulas





true or false at which points in the diagrams?

$b2$, $K_M b3$, $K_J w2$, $K_J (K_M b2 \vee K_M \neg b2)$, $K_P K_J w2$

INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

More Core Topics of Logic:

Finding the Valid Laws of

Reasoning about Knowledge

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Valid Principles

$\models \varphi$ φ is **valid**: φ true in all models at all points

logical consequence $\Phi \models \psi$: validity of implication $\& \Phi \to \psi$

valid

$$K(\varphi \to \psi) \to (K\varphi \to K\psi)$$

$$K(\varphi \wedge \psi) \leftrightarrow (K\varphi \wedge K\psi)$$

$$K\varphi \to \varphi$$

$$K\varphi \to KK\varphi$$

$$\neg K\varphi \to K\neg K\varphi$$

Institute for Logic, Language and Computation

Non-validities

not valid

$K(\varphi \vee \psi) \rightarrow (K\varphi \vee K\psi)$

$K_A K_B \varphi \rightarrow K_B K_A \varphi$

semantic counterexamples

1, p --- 2, ¬p $\varphi = p, \ \psi = \neg p$

1, p -- B -- 2, p -- A -- 3, ¬p $\varphi = p$

in point 1: $K_A K_B p, \neg K_B K_A p$

Institute for Logic, Language and Computation

Multi-S5 Axiom System

1 all valid principles and inference rules of **propositional logic**

e.g., Modus Ponens from $\varphi, \varphi \rightarrow \psi$ infer ψ

2 principles of the **minimal modal logic** (K)

$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$, $\langle\rangle\varphi \leftrightarrow \neg K \neg \varphi$

from φ infer $K\varphi$

3 S5 axioms

$K\varphi \rightarrow \varphi$ (reflexivity) $K\varphi \rightarrow KK\varphi$ (transitivity)

$\neg K\varphi \rightarrow K\neg K\varphi$ (symmetry/euclidity)

Institute for Logic, Language and Computation

Meta-Theory: Soundness and Completeness

Theorem (**Soundness**)

If $\vdash_{S5} \varphi$, then $\models \varphi$

provability implies validity (induction on length of derivations)

use: semantic **counterexamples** to show non-provability

Theorem (**Completeness**)

If $\models \varphi$, then $\vdash_{S5} \varphi$

validity (abstract) implies provability (concrete)

Additional Topics

Reasoning and Computing: Decidability

there is a **decision method** (effective computing procedure)

for determining whether a given formula is valid

equivalent: **satisfiability** problem

given a formula, test if it is true at some point in some model

various decision methods exist

INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

Common Knowledge in a Group

difference **everybody knows ≠ common knowledge**

knowing that others know to any depth of iteration

Muddy Children

After playing outside, two of three children have mud on their foreheads. They can only see the others, so they do not know their own status. (This is an inverse of our card games.) Now their Father says: "At least one of you is dirty". He then asks: "Does anyone know if he is dirty?" Children answer truthfully. As questions and answers repeat, what happens?

discovered independently in sociology, economics, philosophy

used/studied also in logic, linguistics, psychology, ⋯

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Common Knowledge Logic

Difference everybody knows $\mathbf{E_G}\varphi \neq$ common knowledge

knowing that others know to any depth of iteration

$\boldsymbol{M}, s \models C_G\phi$ iff *for all t that are reachable from s by some finite sequence of arbitrary* $\rightarrow_i$ *steps* $(i \in G)$: $\boldsymbol{M}, t \models \phi$

Theorem The complete epistemic logic with common knowledge is axiomatized by the following two principles in addition to the minimal epistemic logic, where E_G is the earlier modality for 'everybody in the group knows':

$C_G\,\phi \leftrightarrow (\phi \wedge E_G C_G\,\phi)$ *Fixed-Point Axiom*

$(\phi \wedge C_G(\phi \rightarrow E_G\,\phi)) \rightarrow C_G\,\phi$ *Induction Axiom*

Outlook, Cognitive Science

reasoning about what others know:

Theory of Mind

growing ability in children

how many levels of iteration can humans achieve?

experiments by logicians, psychologists, game theorists

Issue: **third person perspective** of logical system

first person perspective of agents doing the reasoning

Institute for Logic, Language and Computation

Now Let Us be Precise

About How Update Works

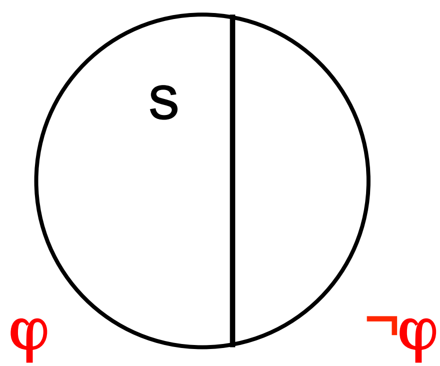
Update with True Information, Picture

epistemic model M, s ~ group information state

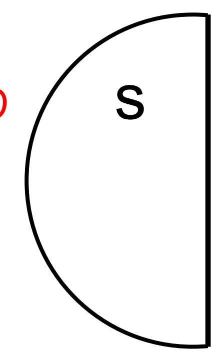
actual world s (seen as such by us as modelers)

!φ learning that φ is true **eliminates all ¬φ-worlds**

from **M**



to submodel **M**|φ



hard information

Institute for Logic, Language and Computation

Typical Dynamic Phenomena

not just announcing facts

communicating ignorance can be informative:

Muddy Children learn facts by learning about their ignorance

typical for dynamics: **order matters**

$!\neg K_{you}\ p\ ;\ !p$ different effect from $!p\ ;\ !\neg K_{you}\ p$

self-refuting information ‘Moore sentences’

$!(\neg Kp \wedge p)$ may give true information,

but you cannot know $\neg Kp \wedge p$

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Extending Epistemic Logic with Dynamic Modalities Expressing the Effects of Information Update

Institute for Logic, Language and Computation

Public Announcement Language & Semantics

simple pilot system for richer dynamic-epistemic logics

PAL **language**

grammar of multi-agent epistemic logic plus **[!φ]ψ**

, $\neg Kp \wedge [!p]Kp$ $[!(\neg K_1 p \wedge [!q]K_2 r)]K_3 \neg K_1 s$

PAL **semantics**

M, s |= **[!φ]ψ** iff if **M**, s |= φ, then **M|**φ, s |= ψ

PAL Axiom System

In technical settings, we often write modal box □ for **K**

axiom system for PAL

1 all proof principles of multi-agent S5

2 all proof principles of basic modal logic for **[!φ]**

plus RE: if $\vdash \varphi \leftrightarrow \psi$, then $\vdash \alpha(\varphi) \leftrightarrow \alpha(\psi)$

3 recursion axioms for postconditions ψ in **[!φ]**ψ:

(a) $[!\varphi]p \leftrightarrow (\varphi \rightarrow p)$ for proposition letters p and the propositional constant T

(b) $[!\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[!\varphi]\psi)$

(c) $[!\varphi](\psi \wedge \alpha) \leftrightarrow ([!\varphi]\psi \wedge [!\varphi]\alpha)$

(d) $[!\varphi]\Box_i\psi \leftrightarrow (\varphi \rightarrow \Box_i(\varphi \rightarrow [!\varphi]\psi))$

(e) $[!\varphi][!\psi]\alpha \leftrightarrow [!(\varphi \wedge [!\varphi]\psi)]\alpha$

Soundness

Theorem All provable formulas of PAL are valid.

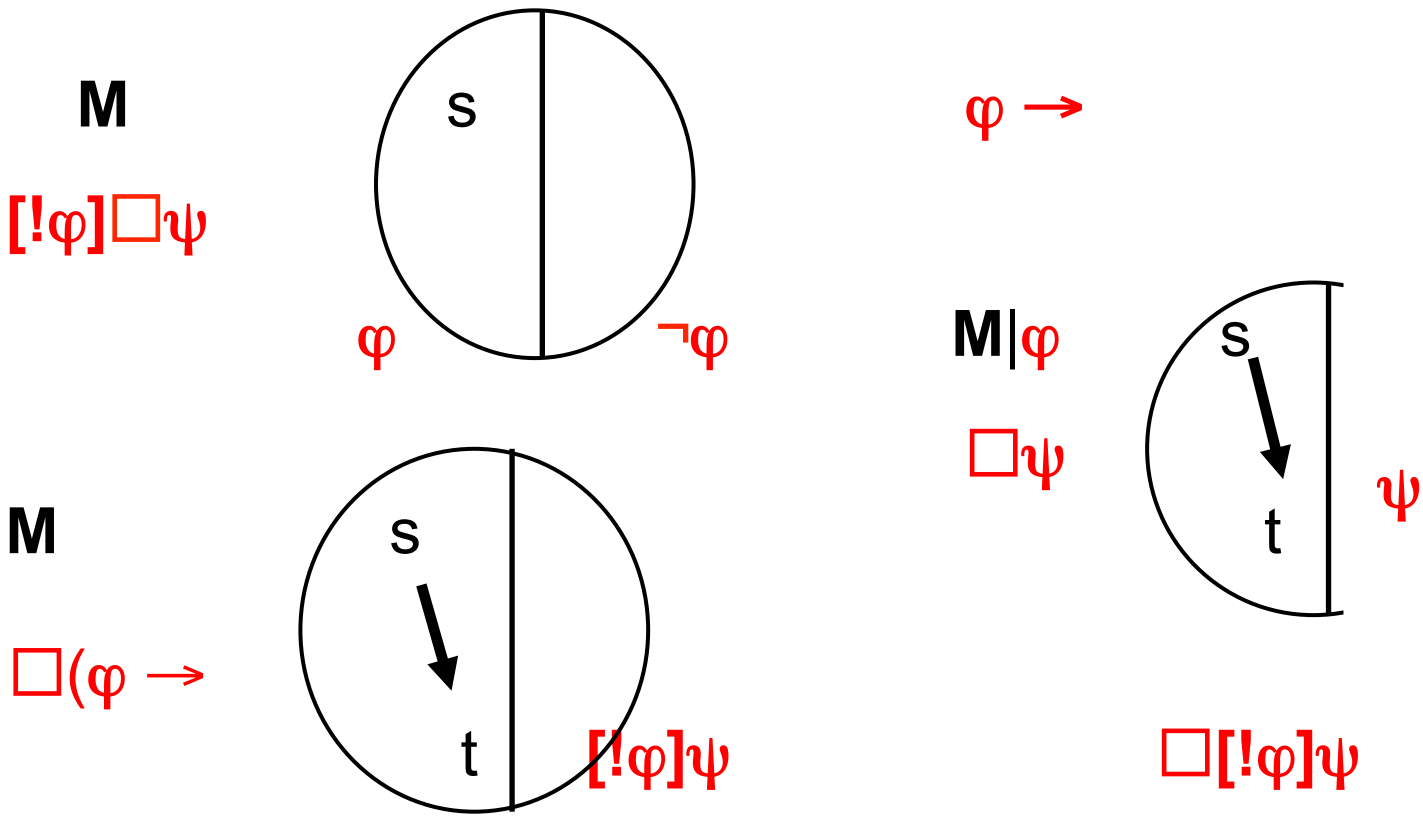
Proof Check what the axioms say

(a) Updates do not affect atomic facts, (b) they are partial functions ($\boldsymbol{M}/\varphi, s$ is unique, but only defined when φ is true at s), (c) is just a modal law, (d) reflects the passage to sub-models, or in a more abstract sense: 'Perfect Recall' and 'No Miracles' properties of PAL agents, (e) several statements can be compressed to a single, more complex, one.

Institute for Logic, Language and Computation

Recursion Axiom for Knowledge After Update

how the stated axiom arises in update diagrams



Institute for Logic, Language and Computation

Non-Valid Principles

easy to give **counterexamples** for the following

with concrete formulas and epistemic models

$[!\varphi]\psi \leftrightarrow (\varphi \rightarrow \psi)$ set $\varphi = p$, $\psi = Kp$

model **1 p -- 2 ¬p**

at world 1: $[!p]Kp$, p, $\neg Kp$, $\neg(p \rightarrow Kp)$ are true

$[!\varphi][!\psi]\alpha \leftrightarrow [!(\varphi \wedge \psi)]\alpha$ set $\varphi = p$, $\psi = K_B q \vee K_B \neg q$, $\alpha = K_A q$

model **1 p,q --A-- 2 p,¬q --B-- 3 ¬p,q**

At 1: $[!p][!(K_B q \vee K_B \neg q)]K_A q$ false, $[!(p \wedge (K_B q \vee K_B \neg q))]K_a q$ true

Completeness

Theorem The provable theorems of PAL are exactly the valid formulas.

Proof Using the recursion axioms each formula can be rewritten step by step to a provably equivalent valid formula in the static epistemic base language. Then use the completeness of the static base logic.

Further Topics

Decidability and Complexity

Theorem PAL is decidable.

The completeness argument also gives a decision procedure.

However, the reduced static formulas may be exponentially larger than the originals.

Theorem The computational complexity of PAL validity and PAL satisfiability is **Pspace-complete**.

The complexity for the epistemic base logic is Pspace-complete, and also, there is another, polynomial reduction for SAT.

PAL with Common Knowledge

difficulty: $[!\varphi]C_G\psi$ has no obvious recursion axiom

new notion: **conditional common knowledge** $C_G^{\alpha}\psi$

at **M**, s, this only looks at reachable points via a path in **M** that runs entirely along points that satisfy α

complete axiomatization

$[!\varphi]C_G\psi \leftrightarrow (\varphi \rightarrow C_G^{\varphi}[!\varphi]\psi)$ valid, but not enough:

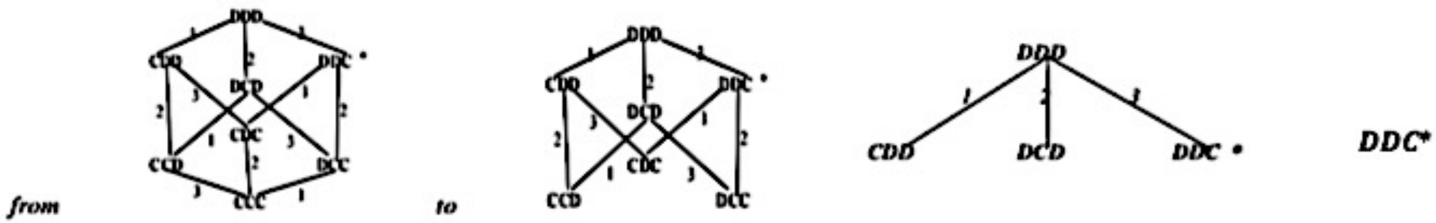
$[!\varphi]C_G^{\alpha}\psi \leftrightarrow (\varphi \rightarrow C_G^{(\varphi \wedge [!\varphi]\alpha)}[!\varphi]\psi)$ works

Muddy Children, PAL Programs

After playing outside, two of three children have mud on their foreheads. They can only see the others, so they do not know their own status. (This is an inverse of our card games.) Now their Father says: "At least one of you is dirty". He then asks: "Does anyone know if he is dirty?" Children answer truthfully. As questions and answers repeat, what happens?

The puzzle is often solved in reasoning, or an ad-hoc mix of reasoning steps and update.

Here is the solution via a sequence of PAL updates:



program structure in communication

WHILE you don't know your status **!**"don't know" ; **!**"I know"

; IF THEN ELSE WHILE DO even **||**

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Outlook, Updates and Inference

real agents mix **update** (observation, communication) and **inference**

知 闻 说 亲 knowledge:

hearing, **proof**, **experience**

Mohist Logic (+– 400 BC) You **see** an object in a dark room and a white object outside the room, someone **tells** you they have the same color, you **infer** the object inside is white.

Does the PAL proof system describe this real behavior?

third person perspective: **theory of information flow**

first person perspective: how agents **themselves** reason