COMP90054 Software Agents
Possible Worlds

Adrian Pearce

Outline

Outline

Equivalence relations

Definition (Kripke Models)

A Kripke model M is a tuple $\langle S, V, R_1, \ldots, R_m \rangle$ where:

Equivalence relations

Definition (Kripke Models)

A Kripke model M is a tuple $\langle S, V, R_1, \ldots, R_m \rangle$ where:

1. S is a non-empty set of *states*, *possible worlds* or *epistemic alternatives*,

Equivalence relations

Definition (Kripke Models)

A Kripke model M is a tuple $\langle S, V, R_1, \ldots, R_m \rangle$ where:

1. S is a non-empty set of *states*, *possible worlds* or *epistemic alternatives*,
2. $V : S \rightarrow (p \rightarrow \{true, false\})$ is a truth assignment to the propositional atoms (p) per state,

Equivalence relations

Definition (Kripke Models)

A Kripke model M is a tuple $\langle S, V, R_1, \ldots, R_m \rangle$ where:

1. S is a non-empty set of *states*, *possible worlds* or *epistemic alternatives*,
2. $V : S \to (p \to \{true, false\})$ is a truth assignment to the propositional atoms (p) per state,
3. $R_i \subseteq S \times S$ (for all $i \in A$) are the *epistemic accessibility* relations for each agent.

Equivalence relations

Definition (Kripke Models)

A Kripke model M is a tuple $\langle S, V, R_1, \ldots, R_m \rangle$ where:

1. S is a non-empty set of *states*, *possible worlds* or *epistemic alternatives*,
2. $V : S \rightarrow (p \rightarrow \{true, false\})$ is a truth assignment to the propositional atoms (p) per state,
3. $R_i \subseteq S \times S$ (for all $i \in A$) are the *epistemic accessibility* relations for each agent.

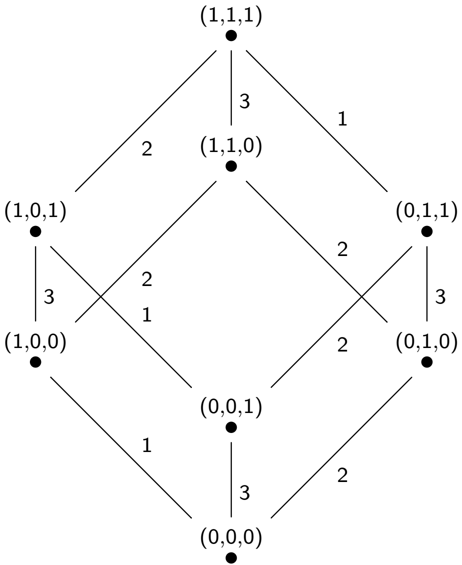
For any state or possible world s,
$(M, s) \models p$ (for $p \in P$) iff $V(s)(p) = $ true

Example: Muddy Children Puzzle

Example

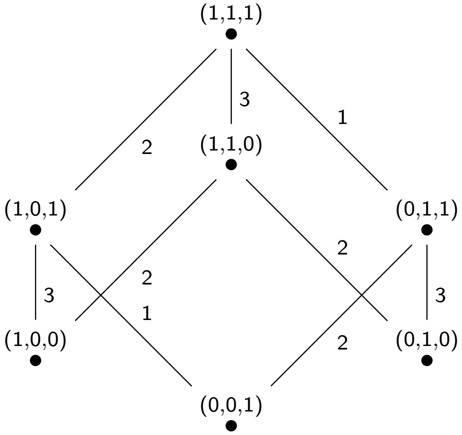
- k children get mud on their foreheads
- Each can see the mud on others, but not on his/her own forehead
- The father says *"at least one of you has mud on your head"* initially.
- The father then repeats *"Can any of you prove you have mud on your head?"* over and over.
- Assuming that the children are perceptive, intelligent, truthful, and that they answer simultaneously, what will happen?

Muddy Children Puzzle (Initially)



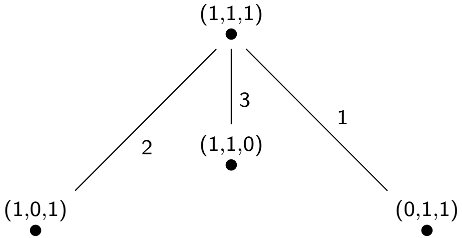
Muddy Children Puzzle (After the father speaks)

Model—general case for all children (Child 1, Child 2, Child 3)



Muddy Children Puzzle (k=1)

First time (k=1) all children say *"No"* and all states with *only one* muddy forehead consequently dissapear.



Muddy Children Puzzle (k=2 & k=3)

Second time (k=2) all children say *"No"* again; this time all states with *only two* muddy foreheads consequently dissapear

(1,1,1)

Third time (k=3) all children say *"Yes"* because they all know their foreheads are muddy (*the model can collapse no further)*.

Forms of knowledge

- $D_G p$: the group G has distributed knowledge of fact p
- $S_G p$: someone in G knows p

$$S_G p \equiv \vee_{i \in G} K_i p$$

- $E_G p$: everyone in G knows p

$$E_G p \equiv \wedge_{i \in G} K_i p$$

Forms of knowledge

- $E_G^k p$ for $k \geq 1$: $E_G^k p$ is defined by

$$E_G^1 p = E_G p$$

$$E_G^{k+1} p = E_G E_G^k p \text{ for } k \geq 1$$

- $C_G p$: p is common knowledge in G

$$C_G \equiv E_G p \wedge E_G^2 p \wedge \ldots E_G^m p \wedge \ldots$$

Synchronisation

Example (The Coordinated Attack Problem (Byzantine Generals))

- Suppose General A sends a message to General B saying *Let's attack at Dawn*.
- Does not have any common knowledge fixpoint (in spite of acknowledgements).
- It seems that common knowledge is theoretically unachievable - how can this be so?

In the presence of unreliable communication, common knowledge is theoretically unachievable.

Simultaneity

In practice, we can establish ϵ-common knowledge, Halpern and Moses (1990).

Definition (ϵ-common knowledge)

ϵ-common knowledge assumes that within an interval ϵ everybody knows ϕ.

Knowledge

Agent i knows p in world s of (Kripke) structure M, exactly if p is true at all worlds that i considers possible in s. Formally,

Knowledge

Agent i knows p in world s of (Kripke) structure M, exactly if p is true at all worlds that i considers possible in s. Formally,

$(M, s) \models K_i p$ iff $(M, t) \models p$ for all t such that $(s, t) \in K_i$

Knowledge

Agent i knows p in world s of (Kripke) structure M, exactly if p is true at all worlds that i considers possible in s. Formally,

$(M, s) \models K_i p$ iff $(M, t) \models p$ for all t such that $(s, t) \in K_i$

Relationship between knowledge forms, D_G, E_G and C_G:

Knowledge

Agent i knows p in world s of (Kripke) structure M, exactly if p is true at all worlds that i considers possible in s. Formally,

$(M, s) \models K_i p$ iff $(M, t) \models p$ for all t such that $(s, t) \in K_i$

Relationship between knowledge forms, D_G, E_G and C_G:

- $\models E_G p \Leftrightarrow \wedge_{i \in G} K_i p$

Knowledge

Agent i knows p in world s of (Kripke) structure M, exactly if p is true at all worlds that i considers possible in s. Formally,

$(M, s) \models K_i p$ iff $(M, t) \models p$ for all t such that $(s, t) \in K_i$

Relationship between knowledge forms, D_G, E_G and C_G:

- $\models E_G p \Leftrightarrow \wedge_{i \in G} K_i p$
- The notions of group knowledge form a hierarchy

$$C_G\varphi \supset \ldots \supset E_G^{k+1}\varphi \supset \ldots \supset E_G\varphi \supset S_G\varphi \supset D_G\varphi \supset \varphi$$

The properties of Knowledge (S5 axioms)

1. $K_i\varphi \wedge K_i(\varphi \Rightarrow \Psi)) \Rightarrow K_i \ \Psi$ (Distribution axiom)
2. *if* $M \models \varphi$ *then* $M \models K_i\varphi$ (Knowledge generalisation rule)
3. $K_i\varphi \Rightarrow \varphi$ (Knowledge or truth axiom)
4. $K_i\varphi \Rightarrow K_iK_i\varphi$ (Positive introspection axiom)
5. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$ (Negative introspection axiom)

Publications

- Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. Rea- soning about Knowledge. The MIT Press, Cambridge, Massachesetts, 1995.
- Joseph Y. Halpern and Yoram Moses, Knowledge and Common Knowledge in a Distributed Environment, Journal of the ACM, Vol. 37, No. 3, pp. 549–587, 1990.
- Richard Scherl and Hector Levesque. Knowledge, Action, and the Frame Problem. Artificial Intelligence, 144:1-39, 2003.

Summary