

**Introductory Course
Epistemic Logic and
Multiagent Systems**

ESSLLI 2003 Vienna

**Hans van Ditmarsch &
Rineke Verbrugge
August 2003**

Lecture 2

Multi-agent epistemic logic

Aims of today's presentation

- Present axiom systems for multi-agent epistemic logics
- Introduce the corresponding Kripke semantics
- Show correspondence between axioms and semantics:
 - characterization of structural properties
 - soundness
 - completeness
- Show how Kripke semantics helps to model distributed systems
- Illustrate this with a case study: communication protocols

Axioms and rules of $K_{(m)}$

(MvdH 1.4; V 1.2)

The basic epistemic logic $K_{(m)}$, where we have an operator K_i for every $i \leq m$, contains the axioms A1, A2, and the derivation rules R1 and R2:

A1 all instances of tautologies of propositional logic

A2 $(K_i\phi \wedge K_i(\phi \rightarrow \psi)) \rightarrow K_i\psi$, for all $i \leq m$

R1 $\vdash \phi, \vdash \phi \rightarrow \psi \Rightarrow \vdash \psi$

R2 $\vdash \phi \Rightarrow \vdash K_i\phi$, for all $i \leq m$

A formula ϕ is *derivable* from $K_{(m)}$, notation $K_{(m)} \vdash \phi$, if there is a proof for ϕ that only uses the axioms and rules of $K_{(m)}$. Sometimes the name of the system is left out, e.g. $\vdash \phi$.

Additional axioms for $T_{(m)}$, $S4_{(m)}$ and $S5_{(m)}$

(MvdH 1.5; V 1.3)

In stronger systems, in addition to $K_{(m)}$, a choice is made from among the following axioms:

A3 $K_i\varphi \rightarrow \varphi$, for all $i \leq m$
known facts are true

A4 $K_i\varphi \rightarrow K_iK_i\varphi$, for all $i \leq m$
positive introspection: an agent knows that it knows something

A5 $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$, for all $i \leq m$
negative introspection: an agent knows that it does not know something

For historical reasons, some of the best-known systems have been given the following names:

$T_{(m)} = K_{(m)} + A3$

$S4_{(m)} = T_{(m)} + A4$

$S5_{(m)} = S4_{(m)} + A5$

Examples of axiomatic proofs

A mixed theorem: $S5_{(2)} \vdash K_2K_1p \rightarrow K_2p$ (Agents know that other agents know only facts)

$$1 \quad S5_{(2)} \vdash K_1p \rightarrow p \quad (\text{A3})$$

$$2 \quad S5_{(2)} \vdash K_2(K_1p \rightarrow p) \quad (\text{R2: 1})$$

$$3 \quad S5_{(2)} \vdash K_2(K_1p \rightarrow p) \rightarrow (K_2K_1p \rightarrow K_2p) \\ (\text{Example 1 in V: 1.2})$$

$$4 \quad S5_{(2)} \vdash K_2K_1p \rightarrow K_2p \quad (\text{R1: 2,3})$$

Another mixed theorem: $S5_{(2)} \vdash K_2K_1p \rightarrow K_2p$ (Agents know that other agents have positive introspection)

$$1 \quad S5_{(2)} \vdash K_1p \rightarrow K_1K_1p \quad (\text{A4})$$

$$2 \quad S5_{(2)} \vdash K_2(K_1p \rightarrow K_1K_1p) \quad (\text{R2: 1})$$

$$3 \quad S5_{(2)} \vdash K_2(K_1p \rightarrow K_1K_1p) \rightarrow (K_2p \rightarrow K_2K_1K_1p) \\ (\text{Example 1 in V: 1.2})$$

$$4 \quad S5_{(2)} \vdash K_2K_1p \rightarrow K_2K_1K_1p \quad (\text{R1: 2,3})$$

Examples of axiomatic proofs, continued

$$1 \text{ S5}_{(m)} \vdash K_1 \neg p \rightarrow \neg p \quad (\text{A3})$$

$$2 \text{ S5}_{(m)} \vdash (K_1 \neg p \rightarrow \neg p) \rightarrow (p \rightarrow \neg K_1 \neg p) \quad (\text{A1})$$

$$3 \text{ S5}_{(m)} \vdash p \rightarrow \neg K_1 \neg p \quad (\text{R1: 1,2})$$

$$4 \text{ S5}_{(m)} \vdash \neg K_1 \neg p \rightarrow K_1 \neg K_1 \neg p \quad (\text{A5})$$

$$5 \text{ S5}_{(m)} \vdash (p \rightarrow \neg K_1 \neg p) \rightarrow \\ ((\neg K_1 \neg p \rightarrow K_1 \neg K_1 \neg p) \rightarrow (p \rightarrow K_1 \neg K_1 \neg p)) \quad (\text{A1})$$

$$6 \text{ S5}_{(m)} \vdash (\neg K_1 \neg p \rightarrow K_1 \neg K_1 \neg p) \rightarrow (p \rightarrow K_1 \neg K_1 \neg p) \\ (\text{R1: 3,5})$$

$$7 \text{ S5}_{(m)} \vdash p \rightarrow K_1 \neg K_1 \neg p \quad (\text{R1: 4,6})$$

Kripke semantics for m agents

(MvdH 1.2 and 1.3; V 1.1)

A Kripke model M is a tuple $M = \langle S, \pi, R_1, \dots, R_m \rangle$ where:

- S is a non-empty set of states s
- π gives, for every state s and atom p , the truth-value $\pi(s)(p)$
- each R_i for $i \leq m$ is a binary accessibility relation between states

We define the relation $(M, s) \models \varphi$, standing for “formula φ is true in state s of model M ”, by induction on the structure of φ :

$$\begin{array}{ll} (M, s) \models p & \text{iff } \pi(s)(p) = \text{true} \\ (M, s) \models (\varphi_1 \wedge \varphi_2) & \text{iff } (M, s) \models \varphi_1 \text{ and } (M, s) \models \varphi_2 \\ (M, s) \models \neg\varphi & \text{iff not } (M, s) \models \varphi \\ (M, s) \models K_i\varphi & \text{iff } \forall t((s, t) \in R_i \Rightarrow (M, t) \models \varphi) \end{array}$$

Example:
modeling a puzzle situation with Kripke semantics

(V 1.1)

The situation: two wise persons, Abélard (A) and Héloïse (H), plus a king

It is general knowledge that:

- there are three hats, two red ones and a white one;
- the king puts one hat each on the heads of A and H;
- A and H cannot see their own hat, but can see the other person's hat;
- the king has asks A whether he knew the color of his own hat, and he answers “no”;
- next, the king asks H whether she know the color of his own hat, and he answers “yes”.

Question: what is the color of Héloïse's hat?

Example, continued

The situation just after the king has put the hats on the two wise persons' heads but before he has asked any questions.

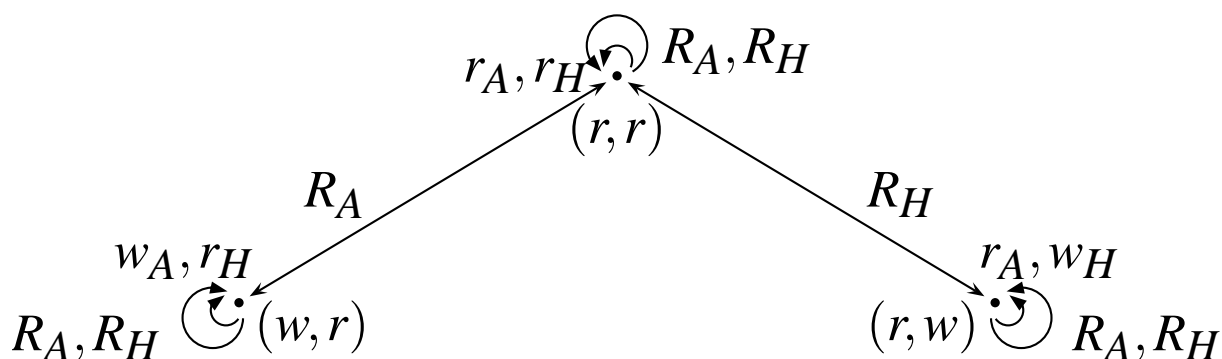
Three states, characterized by the color of the hat of *both* persons. E.g, in (r, w) Abelard wears a red hat and Héloïse a white one.

(w, w) is definitely not a possible situation

Accessibility relations: R_A and R_H

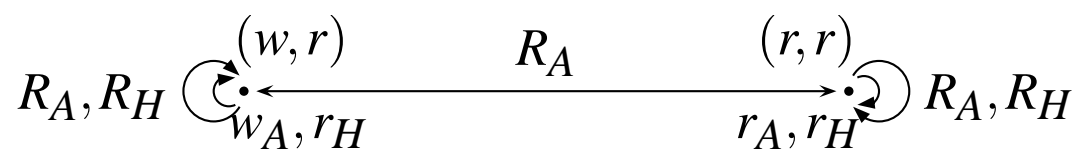
Relevant propositional constants: r_A, r_H, w_A and w_H

In pictures of Kripke models, the convention is to write down only the propositional constants that are *true* in a world.



Example, continued

The situation after the king has asked Abelard whether he knows the color of his hat and Abelard has answered “no”.



Structural properties of Kripke models (MvdH 1.6; V 1.3)

Let S be the set of states of a Kripke model and $R \subseteq S \times S$ an accessibility relation.

<i>R</i> is reflexive	iff $\forall s \in S (s, s) \in R$
<i>R</i> is transitive	iff $\forall s, t, u \in S : (s, t) \in R$ and $(t, u) \in R$ $\Rightarrow (s, u) \in R$
<i>R</i> is symmetric	iff $\forall s, t \in S : (s, t) \in R \Rightarrow (t, s) \in R$
<i>R</i> is euclidean	iff $\forall s, t, u \in S : (s, t) \in R$ and $(s, u) \in R$ $\Rightarrow (t, u) \in R$
<i>R</i> is serial	iff $\forall s \in S \exists t \in S (s, t) \in R$
<i>R</i> is an equivalence relation	iff <i>R</i> is reflexive, transitive and symmetrical

Facts:

R is symmetrical and transitive $\Rightarrow R$ is Euclidean

R is reflexive $\Rightarrow R$ is serial

R is symmetrical, transitive and serial \Leftrightarrow

R is reflexive and euclidean \Leftrightarrow

R is an equivalence relation.

A bit of correspondence theory

(Not in material)

A Kripke frame F is a tuple $F = \langle S, R_1, \dots, R_m \rangle$ (like a Kripke model without valuation).

A formula φ is *valid in model* M , written $M \models \varphi$, if $(M, s) \models \varphi$ for all $s \in S$.

A formula φ is *valid in frame* F , written $F \models \varphi$, if for all valuations π , φ is valid in model $M = \langle S, \pi, R_1, \dots, R_m \rangle$.

A formula is *valid*, written $\models \varphi$, if φ is valid in all models M .

Some example correspondences

For all frames $F = \langle S, R_1, \dots, R_m \rangle$ and $i \leq m$:

- $F \models K_i p \rightarrow p \Leftrightarrow R_i$ is reflexive
- $F \models K_i p \rightarrow K_i K_i p \Leftrightarrow R_i$ is transitive
- $F \models \neg K_i p \rightarrow K_i \neg K_i p \Leftrightarrow R_i$ is euclidean
- $F \models \neg(K_i p \wedge K_i \neg p) \Leftrightarrow R_i$ is serial
- $F \models p \rightarrow K_i \neg K_i \neg p \Leftrightarrow R_i$ is symmetric

Soundness of $K_{(m)}$

(MvdH 1.4; V 1.2)

An axiom system \mathbf{S} is *sound* w.r.t. a class of models \mathcal{M} , if for all formulas φ : $S \vdash \varphi \Rightarrow M \models \varphi$.

Theorem $K_{(m)}$ is sound w.r.t. the class $\mathcal{K}_{(m)}$ of all Kripke models for m agents.

Proof By induction on the length of derivations. All axioms of $K_{(m)}$ are valid, and validity is preserved under the rules:

- A1: for all instances φ of tautologies of propositional logic, $\models \varphi$
- A2: $\models K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi$
- R1: If $\models \varphi$ and $\models \varphi \rightarrow \psi$, then $\models \psi$
- R2: If $\models \varphi$ then $\models K_i\varphi$

Soundness of $T_{(m)}$, $S4_{(m)}$ and $S5_{(m)}$

(MvdH 1.6; V 1.3)

Let M be a class of models.

φ is *valid on* M , written $M \models \varphi$, if for all $M \in M, M \models \varphi$.

Notation:

- $T_{(m)}$ is the class of all reflexive m -Kripke models
- $S4_{(m)}$ is the class of all reflexive, transitive m -Kripke models
- $S5_{(m)}$ is the class of all equivalence m -Kripke models

Theorem:

- $T_{(m)}$ is sound w.r.t. $T_{(m)}$
- $S4_{(m)}$ is sound w.r.t. $S4_{(m)}$
- $S5_{(m)}$ is sound w.r.t. $S5_{(m)}$

Proof Extending the soundness proof for $K_{(m)}$. Use the facts that A3 is valid on $T_{(m)}$, A4 is valid on $S4_{(m)}$ and A5 is valid on $S5_{(m)}$.

Completeness of $K_{(m)}$

(MvdH 1.4 and 1.6; V 1.2 and 1.3)

An axiom system \mathbf{S} is *complete* w.r.t. a class of models M , if for all formulas φ : $M \models \varphi \Rightarrow S \vdash \varphi$.

Theorem $K_{(m)}$ is complete w.r.t. the class $K_{(m)}$ of all Kripke models for m agents.

Proof sketch By contraposition. We will prove: $K_{(m)} \not\models \varphi \Rightarrow$ there is a model M and state s such that $(M, s) \not\models \varphi$.

Step 1: the Lindenbaum lemma

Definition: A set of formulas Φ is *maximally consistent* iff:

- 1) $K_{(m)} \not\models \neg(\varphi_1 \wedge \dots \wedge \varphi_n)$ for any finite $\varphi_1, \dots, \varphi_n \subseteq \Phi$, i.e. Φ is consistent
- 2) For any $\psi \notin \Phi$, there is a finite $\varphi_1, \dots, \varphi_n \subseteq \Phi$ such that $K_{(m)} \vdash \neg(\psi \wedge \varphi_1 \wedge \dots \wedge \varphi_n)$, i.e. $\Phi \cup \psi$ is inconsistent.

Lemma: Every consistent set of formulas can be extended to a maximally consistent set.

Completeness of $K_{(m)}$, continued

Step 2: Construction of the canonical counter-model M

$M = \langle S, \pi, R_1, \dots, R_m \rangle$ where:

- $S = \{s_\Theta \mid \Theta \text{ is a maximally consistent set of formulas} \}$
- $\pi(s_\Theta)(p) = \text{true}$ if $\pi \in \Theta$
 $\pi(s_\Theta)(p) = \text{false}$ if $\pi \notin \Theta$
- $R_i = \{(s_\Theta, s_\Psi \mid \text{for all } \xi : K_i \xi \in \Theta \Rightarrow \xi \in \Psi \}$

Step 3: Show that there is an s such that $(M, s) \not\models \varphi$.

Truth lemma: For every maximally consistent Θ and for all ψ : $(M, s_\Theta) \models \psi \Leftrightarrow \psi \in \Theta$.

Proof: by induction on ψ , using properties of maximally consistent sets and some tricky derivations in $K_{(m)}$. This is the most complex step.

Conclusion: Suppose $\neg\varphi$ is consistent, then it is contained in some maximally consistent set Φ . Now by the truth lemma, $(M, s_\Phi) \not\models \varphi$.

Modeling distributed systems using Kripke semantics

(MvdH 1.8; V 1.5)

Distributed system with processors $1, \dots, m$ interconnected by communication network

Local states s_i for each processor $i \leq m$ are determined by initial state, messages received, internal actions.

Global state of the system $\mathbf{s} = (s_1, \dots, s_m)$.

Associated Kripke model of a distributed system:

$M = \langle S, \pi, R_1, \dots, R_m \rangle$ where:

- $S = \{(s_1, \dots, s_m) \mid \text{for all } i \leq m, s_i \text{ is a local state of processor } i\}$
- π gives, for every global state \mathbf{s} and atom p , the truth-value $\pi(\mathbf{s})(p)$
- $R_i = \{(\mathbf{s}, \mathbf{t}) \mid s_i = t_i\}$ for all $i \leq m$

These are equivalence relations, thus the model $M \in \mathcal{S5}_{(m)}$.

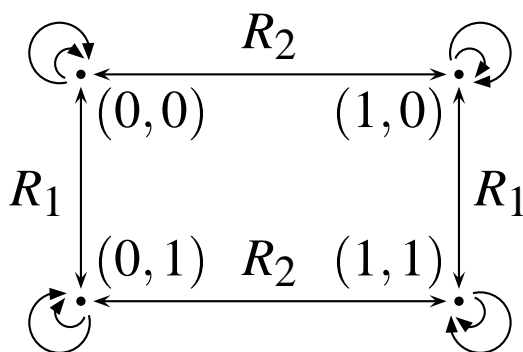
Modeling distributed systems, continued

Example: two processors 1 and 2.

Both of these may be in either of two local states, namely $s_i = 0$ or $s_i = 1$ for $i = 1, 2$. Here s_i refers to the contents of processor i 's register.

The four possible global states: $S = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$.

The accessibility relation is defined according to an informal description of “knowledge”: The processor i “knows” φ iff in every global state which has the same local state as processor i , φ holds, i.e. a processor “only knows itself”.



Case study: investigating communication protocols

(MvdH 1.9; V 1.5)

Two processors: Sender S and Receiver R .

S reads a tape $X = \langle x_0, x_1, \dots \rangle$ with $x_i \in \{0, 1\}$, and sends all inputs it has read to R .

R writes down everything it reads on an output tape Y .

No guarantee that all messages arrive: *deletion errors*.

However, *fairness* holds: if you repeat sending a certain message long enough, an instance of it will eventually arrive.

Sequence transmission problem: can one write a protocol that satisfies the following two constraints, provided that fairness holds?

- *safety*: at any moment, the sequence written by R on Y is a prefix of X
- *liveness*: every $x_i \in X$ will eventually be written by R on Y

Answer: yes. Simulation <http://www.ai.rug.nl/mas/LOK/protocol/>

Investigating communication protocols, continued

$K_S(x_i)$: S knows that the i -th element of X is x_i .

PROTOCOL A FOR S :

```
S1  $i := 0$ 
S2 while true do
S3   begin read  $x_i$ ;
S4     send  $x_i$  until  $K_S K_R(x_i)$ ;
S5     send " $K_S K_R(x_i)$ " until  $K_S K_R K_S K_R(x_i)$ 
S6      $i := i + 1$ 
S7   end
```

PROTOCOL A FOR R :

```
R1 when  $K_R(x_0)$  set  $i := 0$ 
R2 while true do
R3   begin write  $x_i$ ;
R4     send " $K_R(x_i)$ " until  $K_R K_S K_R(x_i)$ ;
R5     send " $K_R K_S K_R(x_i)$ " until  $K_R(x_{i+1})$ 
R6      $i := i + 1$ 
R7   end
```

**Introductory Course
Epistemic Logic and
Multiagent Systems**

ESSLLI 2003 Vienna

**Hans van Ditmarsch &
Rineke Verbrugge
August 2003**

Lecture 4

**Change of common knowledge
over time**

Aims of today's presentation

- Investigate how common knowledge may change
- Illustrate common knowledge change by three case studies:
 1. The muddy children may find out whether they are muddy without being able to see themselves.
 2. Under certain conditions, in distributed systems there is no growth of common knowledge at all.
 3. One can describe (common) knowledge changes in card games like Cluedo

The muddy children puzzle

(MvdH 2.1; V 2.1)

Principal players: a father and n children in a circle around him, of whom k (with $1 \leq k \leq n$) have mud on their forehead.

The following is common knowledge among the children:

- No child knows whether it is muddy or not.
- All of them can accurately perceive whether the other children are muddy.
- All of them are perfect logical reasoners and have successfully finished the ESSLLI course on epistemic logic.
- All of them are perfectly honest and do not cheat.

Father announces:

At least one of you is muddy. If you know that you are muddy, please come forward.

After this, nothing happens. When the father notices this, he literally repeats the announcement. Once again, nothing happens. The announcement and subsequent silence are repeated, but . . .

The muddy children puzzle, analysis

After the father's k -th announcement, suddenly all k muddy children step forward!

Proof Using the epistemic language (E and C w.r.t. n agents):

$\varphi_j =$ there are at least j muddy children

$\psi_j =$ there are precisely j muddy children

We first prove by induction on j ($1 \leq j \leq k$) that after father's j -th questioning round, $C\varphi_j$.

Base case $k = 1$ and father just announced φ_1 , so $C\varphi_1$.

Induction step Suppose $j < k$. Induction hypothesis: After the j -th round, $C\varphi_j$ (and thus $E\varphi_j$).

In the j -th round each muddy child sees only $k - 1$ muddy children, thus cannot conclude that it is muddy itself, and does not step forward.

In the $j + 1$ -st round, all children know that in the j -th round nobody stepped forward (and why), so $C\neg\psi_j$, thus $C\varphi_{j+1}$.

Conclusion In the k -th round, $E\varphi_k$ and, because all muddy children see only $k - 1$ muddy children, they know they are muddy and step forward; the others do not.

The muddy children puzzle, semantics

For n children:

Propositional atoms $\{p_1, \dots, p_n\}$ with
 $p_i = i$ is muddy.

Kripke model for situation just before father's first announcement:

$M = \langle S, \pi, R_1, \dots, R_n \rangle$ with

- $S = \{s = \langle s_1, \dots, s_n \rangle \mid s_i \in \{0, 1\}\}$
- $\pi(\langle s_1, \dots, s_n \rangle)(p_i) = \text{true} \Leftrightarrow s_i = 1$
- $R_i(s, s') \Leftrightarrow s_j = s'_j$ for all $j \neq i$

Given that each child sees the others but not itself, agent 1 can not distinguish between two states $(0, y, z)$ and $(1, y, z)$.

The muddy children puzzle, semantics continued

Example for 3 children: state $(1, 1, 0)$ denotes that child 1 and 2 are muddy, and 3 is not.

$$(M, (1, 1, 0)) \models K_1(p_1 \rightarrow (K_2p_1 \wedge K_2K_3p_1))$$

1 knows that, if he is muddy, then 2 knows this, and also knows that 3 knows it.

However many children are muddy, $M \models \neg C\phi_1$: there is no common knowledge that there is at least one muddy child before father makes his first announcement!

Common knowledge in distributed systems

(MvdH 2.2; V 2.3)

Reminder Modeling distributed systems:

$M = \langle S, \pi, R_1, \dots, R_m \rangle$, where

- $S = \{(s_1, \dots, s_m) \mid \text{for all } i \leq m, s_i \text{ is a local state of processor } i\}$
- $R_i = \{(\mathbf{s}, \mathbf{t}) \mid s_i = t_i\}$ for all $i \leq m$

Such a model does not describe actual state transformations. Epistemic logic can say something about processors' knowledge change during a run.

A *run* in a Kripke model M , with set of states S , is a sequence of states $s^{(1)}, s^{(2)}, \dots$, denoted as $s^{(1)} \rightarrow s^{(2)} \rightarrow \dots$

Warning: \rightarrow has nothing to do with (unions of) accessibility relations.

Question Is it possible to devise a distributed system such that common knowledge *increases* during some run? Thus,

$$(M, s^{(i)}) \not\models C\phi \text{ for } 1 \leq i \leq k$$

$$(M, s^{(i)}) \models C\phi \text{ for } i > k$$

Common knowledge in distributed systems, continued

Answer Under the definition of accessibility from MvdH 1.8:
No, common knowledge does not increase.

A Kripke model is *strongly connected* iff for all $s, t \in S$, $s \twoheadrightarrow t$.

Proposition In $M = \langle S, \pi, R_1, \dots, R_m \rangle$ with $m > 1$:
for all $\mathbf{s}, \mathbf{t} \in S$, there is \mathbf{u} in S such that $R_1(\mathbf{s}, \mathbf{u})$ and $R_2(\mathbf{u}, \mathbf{t})$;
i.e. M is strongly connected.

Proposition In all Kripke models N :
If $(N, s) \models C\varphi$, then for all t with $s \twoheadrightarrow t$, also $(N, t) \models C\varphi$.

Corollary In $M = \langle S, \pi, R_1, \dots, R_m \rangle$ with $m > 1$:
for all $\mathbf{s}, \mathbf{t} \in S$, if $(M, \mathbf{s}) \models C\varphi$, then $(M, \mathbf{t}) \models C\varphi$, i.e. common
knowledge is constant through every run on M .

Common knowledge in distributed systems, continued

More generally:

A run $s^{(1)} \rightarrow s^{(2)} \rightarrow \dots$ is *non-simultaneous* iff for every transition $s^{(k)} \rightarrow s^{(k+1)}$ there exists a processor i , $1 \leq i \leq m$, such that $s_i^{(k)} \rightarrow s_i^{(k+1)}$.

Theorem (Halpern and Moses) In non-simultaneous runs, common knowledge is constant.

These are paradoxical results: in distributed systems, we expect common knowledge to increase by communication.

Solution: in practice, it may be that not all global states of the system are reachable.

Exercise Give an example of a distributed system with $S \subset S_1 \times \dots \times S_m$, together with a run along which common knowledge changes.

Co-ordinated attack is impossible

(MvdH 2.2; V 2.3)

Two allied generals, A en B , stand on two mountain summits, with their enemy in the valley between them.

It is generally known that A and B together can easily beat the enemy, but if only one of them attacks, he will certainly lose the battle.

It is not guaranteed that a messenger will arrive from A to B or vice versa.

A sends messenger to B with the message $m =$ "I propose that we attack Friday at 8 AM sharp".

Suppose the messenger reaches B .

Then $K_B m$, and $K_B K_A m$.

Will it be a good idea to attack?

No, because $\neg K_A K_B m$. Thus, B sends an acknowledgement to A , etcetera.

Co-ordinated attack is impossible, continued

Common knowledge will never be established, because for every $n \geq 0$:

odd rounds After the messenger has safely brought $2n + 1$ messages, $K_B(K_A K_B)^n(m)$ is reached, but $(K_A K_B)^{n+1}(m)$ does not hold.

even rounds After the messenger has safely brought $2n + 2$ messages, $(K_A K_B)^{n+1}(m)$ is reached, but $K_B(K_A K_B)^{n+1}(m)$ does not hold.

This is similar to situation for file transmission protocols like protocol A and TCP.

Theorem (Halpern and Moses) In order to start a coordinated attack, Common Knowledge of m is necessary.

-R. Fagin, J.Y. Halpern, Y.O. Moses and M.Y. Vardi, *Reasoning about Knowledge*, Cambridge (MA) MIT Press, 1995.

-F. Stulp and R. Verbrugge, A knowledge-based algorithm for the Internet protocol TCP, *Bulletin of Economic Research* 54 (1)(2002), 69–94.

Co-ordinated attack is impossible, variation

In the variation, the circumstances are a bit more favorable.

Two parties, S and R , know that their communication channel is trustworthy, but with one catch:

When a message m is sent at time t , it either arrives immediately, or at time $t + \varepsilon$.

This catch is common knowledge between S and R .

Now S sends a message b to R at time t_0 .

Question: When will Common Knowledge about b be established?

Answer: "Never!"

Exercise: Give an analysis in terms of distributed systems.

W. van der Hoek and R. Verbrugge, Epistemic logic: a survey. In: L.A. Petrosjan, V.V. Mazalov (eds.), *Game Theory and Applications*, vol. 8, New York, Nova Science Publishers, 2002, pp. 53-94.

**Introductory Course
Epistemic Logic and
Multiagent Systems**

ESSLLI 2003 Vienna

**Hans van Ditmarsch &
Rineke Verbrugge
August 2003**

Lecture 5

**Belief and
boundaries of epistemic logic**

Aims of today's presentation

- Introduce a weaker notion of group knowledge: distributed knowledge
 - Axiom system
 - Kripke semantics
 - Soundness and completeness

- Investigate the notion of belief
 - Axiom system
 - Kripke semantics
 - Soundness and completeness
 - Group belief

Aims of today's presentation, continued

- Discuss relations between knowledge and belief
 - Can one define one in terms of the other?
 - Default rules in terms of knowledge and preference
- Discuss boundaries of epistemic logic
 - Bounded rationality and the problem of logical omniscience

Distributed knowledge

(MvdH 2.3; V 2.4)

Intuition:

A group of agents has *distributed* or *implicit knowledge* of a fact φ ($I\varphi$) if the knowledge of φ is distributed over the members of that group.

Example:

φ = “Two participants have the same birthday”

Then either $\varphi \wedge I\varphi$ or $\neg\varphi \wedge I\neg\varphi$.

Semantics Given a Kripke model $M = \langle S, \pi, R_1, \dots, R_m \rangle$:

$$(M, s) \models I\varphi \Leftrightarrow (M, t) \models \varphi \text{ for all } t \text{ such that } (s, t) \in R_1 \cap \dots \cap R_m.$$

Distributed knowledge, continued

Dually to the semantic definition:

$$(M, s) \models \neg I\neg\varphi \Leftrightarrow \text{there is a } t \text{ such that} \\ (s, t) \in R_1 \cap \dots \cap R_m \text{ and } (M, t) \models \varphi$$

Thus, if all agents could pool their knowledge, they would only consider those worlds as possible that are epistemic alternatives for *each* of them.

$I\varphi =$ “a wise person knows φ ”

$C\varphi =$ “any fool knows φ ”

Comparing the strengths of different notions of knowledge (over $S5_n$):

$$C\varphi \Rightarrow E\varphi \Rightarrow K_i\varphi \Rightarrow I\varphi \Rightarrow \varphi$$

Distributed knowledge: axioms

A1' all instances of tautologies of propositional logic

A11 $K_i\phi \rightarrow I\phi$, for all $i \leq m$
individual knowledge implies distributed knowledge

A12 $(I\phi \wedge I(\phi \rightarrow \psi)) \rightarrow I\psi$ A2 for I

A13 $D\phi \rightarrow \phi$ A3 for I

A14 $I\phi \rightarrow II\phi$ A4 for I

A15 $\neg I\phi \rightarrow I\neg I\phi$ A5 for I

We introduce the following axiom systems:

$KI_{(m)} = K_{(m)} + A11 + A1' + A12$

$TI_{(m)} = KI_{(m)} + A3 + A13$

$S4I_{(m)} = TI_{(m)} + A4 + A14$

$S5I_{(m)} = S4I_{(m)} + A5 + A15$

Fact: If $KI_{(m)} \vdash (\psi_1 \wedge \dots \wedge \psi_m) \rightarrow \phi$, then

$KI_{(m)} \vdash (K_1\psi_1 \wedge \dots \wedge K_m\psi_m) \rightarrow I\phi$ (Rule R4)

Conversely, A11 follows from $K_{(m)} + A1' + A12 + R4$.

Distributed knowledge: soundness and completeness

Correspondence:

For all frames $F = \langle S, R_1, \dots, R_m \rangle$ and $i \leq m$,

$$F \models K_i p \rightarrow Ip \Leftrightarrow R_I \subseteq R_i$$

Thus, A11 is sound w.r.t. those Kripke models in which the accessibility relation $R_I = R_1 \cap \dots \cap R_m$.

For the semantic class $K_{(m)}$ we now assume that $R_I = R_1 \cap \dots \cap R_m$, similarly for the others.

Theorem Soundness and completeness

$KI_{(m)}$	is sound and complete w.r.t. $K_{(m)}$
$TI_{(m)}$	is sound and complete w.r.t. $T_{(m)}$
$S4I_{(m)}$	is sound and complete w.r.t. $S4_{(m)}$
$S5I_{(m)}$	is sound and complete w.r.t. $S5_{(m)}$

Belief

(MvdH 2.4)

The standard system for doxastic logic is $KD45_{(m)}$, with operator B_i for every $i \leq m$; containing the axioms A1, A2, D, A4 and A5, and the derivation rules R1 and R2:

A1 all instances of tautologies of propositional logic

A2 $(B_i\phi \wedge B_i(\phi \rightarrow \psi)) \rightarrow B_i\psi$, for all $i \leq m$

D $\neg B_i(\perp)$

an agent's belief set is not inconsistent

A4 $B_i\phi \rightarrow B_iB_i\phi$, for all $i \leq m$

positive introspection

A5 $\neg B_i\phi \rightarrow B_i\neg B_i\phi$, for all $i \leq m$

negative introspection

R1 $\vdash \phi, \vdash \phi \rightarrow \psi \Rightarrow \vdash \psi$

R2 $\vdash \phi \Rightarrow \vdash B_i\phi$, for all $i \leq m$

Facts

A3 is not included, because agents may have false beliefs.

D is equivalent in $K_{(m)}$ to $\neg(B_i p \wedge B_i \neg p)$, and follows from A3.

Belief: soundness and completeness

Semantics for belief.

A Kripke model M is a tuple $M = \langle S, \pi, R_1, \dots, R_m \rangle$

$(M, s) \models B_i \varphi$ iff $\forall t ((s, t) \in R_i \Rightarrow (M, t) \models \varphi)$

Notation:

$KD45_{(m)}$ is the class of serial, transitive, euclidean m -Kripke models.

Theorem

$KD45_{(m)}$ is sound and complete w.r.t. $KD45_{(m)}$.

Proof sketch

Canonical model construction as for $K_{(m)}$. Because $\neg B_i(\perp)$ is contained in every maximal $KD45_{(m)}$ -consistent set, the canonical model is serial. Similarly, A4 makes the model transitive and A5 makes it euclidean.

Collective belief

(Not in material)

Notation:

$E_B p$: everybody believes p

$C_B p$: the group collectively believes p

Semantics:

$(M, s) \models C_B \varphi$ iff

$\forall t$ (t is “reachable” in 1 or more steps from $s \Rightarrow (M, t) \models \varphi$)

Axiom system:

$KD45_{(m)}^{EC}$ is $KD45_{(m)}$ plus axioms and rule:

A6' $E_B \varphi \leftrightarrow B_1 \varphi \wedge \dots \wedge B_m \varphi$

A8' $C_B \varphi \rightarrow E_B C_B \varphi$

A9' $(C_B \varphi \wedge C_B(\varphi \rightarrow \psi)) \rightarrow C_B \psi$

A10' $C_B(\varphi \rightarrow E_B \varphi) \rightarrow (E \varphi \rightarrow C_B \varphi)$

R3' If $\vdash \varphi$, then $\vdash C_B \varphi$

$C_B \varphi \rightarrow \varphi$ does *not* hold: group may have collective illusion.

Theorem:

$KD45_{(m)}^{EC}$ is sound and complete w.r.t. $KD45_{(m)}$.

Relations between knowledge and belief

(Not in material, but see MvdH 2.13)

Example:

Muddy children with 3 children A,B,C, all muddy.

Old assumptions hold, but A and B cheat: before round 1, they signal “you are muddy”. C cannot imagine such secret communication.

In round 2, A and B answer “yes, I know I’m muddy”.

C believes but does not know that she is not muddy.

If C’s trustfulness is common knowledge, then it is common knowledge that C believes she is muddy.

Variations:

But what if C suspects something?

What if she has actually (she believes secretly) observed the communication between A and B, while they think this is impossible?

A. Baltag, A logic for suspicious players: epistemic actions and belief-updates in games, <http://www.cwi.nl/~abaltag>

Relations between knowledge and belief, continued

Philosophers: knowledge is

- justified
- true
- belief

But even this may not be strong enough!

E. Gettier, Is justified true belief knowledge?, *Analysis* 23 (1963), pp. 121-123.

In literature on multi-agent systems, often the following axiom is used:

$$K_i\varphi \leftrightarrow (\varphi \wedge B_i\varphi)$$

Clearly, a true belief may not be *justified*, if the agent believes it for the wrong reason.

Example of modified muddy children: C is not muddy. After round 2, does she *know* she's not?

Relations between knowledge and belief, continued

Kraus and Lehmann have a more complex system $KL_{(m)}$ combining knowledge and belief.

S. Kraus and D. Lehmann, Knowledge, belief, and time, *Theoretical Computer Science* 58 (1988), 155-174.

$KL_{(m)}$ is $S5_{(m)}$ for the K-operators and $KD45_{(m)}$ for the B-operators, plus:

$$\text{KB1} \quad K_i\varphi \rightarrow B_i\varphi$$

$$\text{KB2} \quad B_i\varphi \rightarrow K_iB_i\varphi$$

Problem

It appears that in $KL_{(m)}$, an agent cannot believe to know a false proposition!

$$KL_{(m)} \vdash B_iK_i\varphi \rightarrow K_i\varphi$$

Relations between knowledge and belief, continued

Proof sketch for $KL_{(2)} \vdash B_1K_1p \rightarrow K_1p$

- 1 $KL_{(2)} \vdash B_1K_1p \rightarrow \neg B_1\neg K_1p$
(D in its form $\neg(B_i\phi \wedge B_i\neg\phi)$ + prop. logic)
- 2 $KL_{(2)} \vdash \neg B_1\neg K_1p \rightarrow \neg K_1\neg K_1p$
(KB1 and prop. logic: contraposition)
- 3 $KL_{(2)} \vdash \neg K_1\neg K_1p \rightarrow K_1p$
(A5 and prop. logic: contraposition)
- 4 $KL_{(2)} \vdash B_1K_1p \rightarrow K_1p$
(from 1,2,3 by prop. logic: hypothetical syllogism)

For more on combining knowledge and belief, see
F. Voorbraak, The logic of objective knowledge and rational
belief, in J. van Eijck (ed.), *Logics in AI*, LNCS 478, Berlin,
Springer, 1991, pp. 499-515.

Default logic in terms of knowledge

(Not in material, but see MvdH Ch. 4)

A default (Reiter) with φ, ψ, χ propositional formulas:

$$\frac{\varphi : \psi}{\chi}$$

Intuitively:

“If you know φ , and ψ is consistent with your knowledge, then add χ to your knowledge (but check whether ψ remains consistent with your current knowledge)”.

Example:

$$\frac{\text{esslli}(k) : \text{likes} - \text{logic}(k)}{\text{likes} - \text{logic}(k)}$$

Default logic in terms of knowledge

Two ways to represent these defaults

$$\frac{\varphi : \psi}{\chi}$$

in 1-agent epistemic logic with $M\varphi = \neg K\neg\varphi$ and with extra operator P (preferred):

1 strong representation $\varphi \wedge M\psi \rightarrow P\chi$

If φ is true and ψ is considered possible, then χ is preferred

2 weak representation $K\varphi \wedge M\psi \rightarrow P\chi$

If φ is known and ψ is considered possible, then χ is preferred

Default logic in terms of knowledge, continued

Language: propositional plus modal operators K, M, P_1, \dots, P_n

Semantics: A Kripke model $M \in S5P$ is a tuple $M = \langle S, \pi, S_1, \dots, S_n, P, P_1, \dots, P_n \rangle$ where:

- $S_i \subseteq S$ are sets ('frames of reference') of preferred states
- $P = S \times S$ (so is an equivalence relation)
- $P_i = S \times S_i$ (so is transitive and Euclidean)

We define the relation $(M, s) \models \varphi$:

$(M, s) \models p$	iff	$\pi(s)(p) = true$
$(M, s) \models (\varphi_1 \wedge \varphi_2)$	iff	$(M, s) \models \varphi_1$ and $(M, s) \models \varphi_2$
$(M, s) \models \neg\varphi$	iff	not $(M, s) \models \varphi$
$(M, s) \models K\varphi$	iff	$\forall t((s, t) \in P \Rightarrow (M, t) \models \varphi)$
$(M, s) \models M\varphi$	iff	$\exists t((s, t) \in P$ and $(M, t) \models \varphi)$
$(M, s) \models P_i\varphi$	iff	$\forall t((s, t) \in P_i \Rightarrow (M, t) \models \varphi)$

$\forall s, t((s, t) \in P_i \Leftrightarrow t \in S_i)$.

Thus $P_i\varphi$ (" φ is a practical belief w.r.t. frame of reference S_i ") is true if φ is true in all states of S_i .

Default logic in terms of knowledge, continued

Example: Bush diamond

Take the default theory:

$$1) \frac{r : \neg p}{\neg p}$$

$$2) \frac{m : p}{p}$$

$$3) r \wedge m$$

- r Bush is a Republican
- p Bush is a pacifist
- m Bush is a Methodist

In the strong EDL-representation 1) and 2) become:

$$1) r \wedge M\neg p \rightarrow \neg p$$

$$2) m \wedge Mp \rightarrow p$$

Suppose $Mp \wedge M\neg p$, then $P_1\neg p$ and P_2p .

Two disjunct subframes S_1 (where $\neg p$) and S_2 (where p).

Cf. Reiter extensions $Th(\{r, m, p\})$ and $Th(\{r, m, \neg p\})$.

Two modalities P_1 and P_2 needed:
one modality causes empty frame.

Default logic in terms of knowledge, continued

Axiom system S5P contains:

S5 for K-operator (with dual M)

K45 for P-operator, plus:

$$\text{AP6} \quad K\varphi \rightarrow P_i\varphi$$

(“certain” implies “preferred” in any frame of reference)

$$\text{AP7} \quad P_i\varphi \leftrightarrow K_iP_i\varphi$$

$$\text{AP8} \quad \neg P_i\perp \rightarrow (P_iP_j\varphi \leftrightarrow P_j\varphi)$$

$$\text{AP9} \quad \neg P_i\perp \rightarrow (P_iK\varphi \leftrightarrow K\varphi)$$

Nested modalities are reduced to innermost one.

In AP8 and AP9, condition $\neg P_i\perp$ is needed to prevent left-hand side of equivalences to be trivially true.

Correspondences:

AP6 the P_i are sub-relations of P

AP7 the P_i reach same S_i , whatever starting point in S

AP8 the P_j reach same S_j , independent of starting point in some non-empty frame S_i

AP9 P reach whole S , independent of starting point in some non-empty frame S_i

Theorem S5P is sound and complete w.r.t. $S5P$

The logical omniscience problem

(MvdH 2.5)

Problems of any standard modal approach (extending K) to belief and knowledge:

- | | | |
|---|---|---------------------------------|
| 1 | $K\phi \wedge K(\phi \rightarrow \psi) \rightarrow K\psi$ | closure under implication |
| 2 | $\models \phi \Rightarrow \models K\phi$ | knowledge of valid formulas |
| 3 | $\models \phi \rightarrow \psi \Rightarrow \models K\phi \rightarrow K\psi$ | closure under valid implication |
| 4 | $\models \phi \leftrightarrow \psi \Rightarrow \models K\phi \leftrightarrow K\psi$ | knowledge of equivalent form. |
| 5 | $(K\phi \wedge K\psi) \rightarrow K(\phi \wedge \psi)$ | closure under conjunction |
| 6 | $K\phi \rightarrow K(\phi \vee \psi)$ | weakening of knowledge |
| 9 | $K(p \vee \neg p)$ | knowledge of truth |

for systems extending D and A5, moreover:

- | | | |
|---|------------------------------------|-----------------------------------|
| 7 | $B\phi \rightarrow \neg B\neg\phi$ | consistency of beliefs |
| 8 | $B(B\phi \rightarrow \phi)$ | belief of having no false beliefs |

Exercise: Find real-life counterexamples showing that properties 1-6 and 9 do not model human knowledge / belief and 7-8 do not model belief.

The logical omniscience problem, continued

(not in material, but see MvdH 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, 2.12)

One of many proposed solutions to the logical omniscience problem:

Fagin and Halpern: Awareness and explicit belief

Language: Awareness A_i and explicit belief $B_{E,i}$ added.

Kripke model: $M = \langle S, \pi, A_1, \dots, A_n, R_1, \dots, R_n \rangle$

- the R_i ($i \leq n$) are serial, transitive and Euclidean
- A_i is a function that yields, for any $s \in S$, the set of formulas of which agent i is aware

Semantics as usual, but:

$$(M, s) \models A_i\varphi \Leftrightarrow \varphi \in A_i(s)$$

$$(M, s) \models B_{E,i}(\varphi) \Leftrightarrow \varphi \in A_i(s) \text{ and } (M, t) \models \varphi \text{ for all } t \text{ such that } (s, t) \in R_i.$$

Thus, $B_{E,i}\varphi \leftrightarrow (A_i\varphi \wedge B_i\varphi)$ holds.

The logical omniscience problem, continued

Examples:

$\{B_{E,i}p, B_{E,i}(p \rightarrow q), \neg B_{E,i}q\}$ is satisfiable (cf. 1)

$\{\neg B_{E,i}(p \vee \neg p)\}$ is satisfiable (cf. 9)

$\{B_{E,i}(p \wedge \neg p)\}$ is *not* satisfiable (cf. 7)

Problem with FH approach:

By choosing suitable functions A_i , one can deny even the intuitive properties:

$$A_i(p \wedge q) \rightarrow A_i(q \wedge p)$$

$$B_{E,i}p \rightarrow B_{E,i}B_{E,i}p$$

$$\neg B_{E,i}p \rightarrow B_{E,i}\neg B_{E,i}p.$$

R. Fagin and J.Y. Halpern, Belief, awareness, and limited reasoning, *Artificial Intelligence* 34 (1988), pp. 39-76.

The logical omniscience problem between agents

(not in material, but see MvdH 2.9)

Transparency: $S5_{(2)} \vdash K_2 K_1 p \rightarrow K_2 p$

If child 2 knows that her father 1 has proved Fermat's Last Theorem, then does the child know the theorem?

Solution by Gochet and Gillet: Local worlds

Agent 2 cannot consult the alternatives that agent 1 considers as compatible with 2's alternatives of the real world, e.g. agent 2 does not know whether R_1 is reflexive, i.e. whether 1's knowledge is true.

Kripke model: $M = \langle S_1, \dots, S_n, \rho, \pi, R_1, \dots, R_n \rangle$

- S_i is the set of states associated with agent i
- the R_i ($i \leq n$) are equivalence relations
- $\rho \in \bigcap_i S_i$ is the "real" state

The logical omniscience problem between agents, cont.

Semantics two types of validity:

a formula is valid₁ iff for every model M and any $s \in \bigcup_i S_i$,
 $(M, s) \models \varphi$

a formula is valid₂ iff for every model M , $(M, \rho) \models \varphi$

Examples:

$K_1(p \rightarrow q) \rightarrow (K_1p \rightarrow K_1q)$	valid ₁ and valid ₂
$K_1p \rightarrow p$	valid ₂ , not valid ₁
$K_2K_1p \rightarrow K_2p$	not valid ₁ , not valid ₂

P. Gochet and E. Gillet, On Professor Weingartner's contribution to epistemic logic. In G. Schurz and G.J.W. Dorn (eds.), *Advances in Scientific Philosophy*, Amsterdam, Rodopi, 1991, pp. 97-115.