Modal Logics for Multi-Agent Systems

Valentin Goranko[1] and Wojtek Jamroga[2]

[1] University of the Witwatersrand, Johannesburg, South Africa

[2] Clausthal University of Technology, Germany

Time:

31 July–4 August 2006, 17:00–18:30

Organization:

Lectures 1, 3, 5, 8 & 10: Wojtek Jamroga,
Lectures 2, 4, 6, 7, & 9: Valentin Goranko.

More course material, including updated notes, exercises, slides, and references will be placed on the course website:

http://www2.in.tu-clausthal.de/∼wjamroga/courses/MAS2006ESSLLI/

Section 1. Agents and Logics

1.1 Agents

- Multi-agent system (MAS): a system that involves several autonomous entities that act in the same environment
- The entities are called agents

- Multi-agent system (MAS): a system that involves several autonomous entities that act in the same environment
- The entities are called agents

- So, what is an agent precisely?

- Multi-agent system (MAS): a system that involves several autonomous entities that act in the same environment
- The entities are called agents

- So, what is an agent precisely?
- No commonly accepted definition

For some authors, agents are:

- A new paradigm for computation

For some authors, agents are:

- A new paradigm for computation
- A new paradigm for design

For some authors, agents are:

- A new paradigm for computation
- A new paradigm for design
- A new paradigm for programming

For some authors, agents are:

- A new paradigm for computation
- A new paradigm for design
- A new paradigm for programming

Our claim:

MAS is a philosophical metaphor that induces a specific way of seeing the world.

Intuition:

We are agents!

Intuition:

We are agents!

The metaphor:

- Makes us use specific vocabulary

Intuition:

We are agents!

The metaphor:

- Makes us use specific vocabulary
- Makes us use specific conceptual structures

Intuition:

We are agents!

The metaphor:

- Makes us use specific vocabulary
- Makes us use specific conceptual structures
- So:
- A new paradigm for thinking and talking about the world

Features of agents

An agents can/should possibly be:

Features of agents

An agents can/should possibly be:

- Autonomous: operates without direct intervention of others, has some kind of control over its actions and internal state

Features of agents

An agents can/should possibly be:

- Autonomous: operates without direct intervention of others, has some kind of control over its actions and internal state
- Reactive: reacts to changes in the environment

Features of agents

An agents can/should possibly be:

- Autonomous: operates without direct intervention of others, has some kind of control over its actions and internal state
- Reactive: reacts to changes in the environment
- Pro-active: takes the initiative

Features of agents

An agents can/should possibly be:

- Autonomous: operates without direct intervention of others, has some kind of control over its actions and internal state
- Reactive: reacts to changes in the environment
- Pro-active: takes the initiative
- Goal-directed: acts to achieve a goal

Features of agents

An agents can/should possibly be:

- Autonomous: operates without direct intervention of others, has some kind of control over its actions and internal state
- Reactive: reacts to changes in the environment
- Pro-active: takes the initiative
- Goal-directed: acts to achieve a goal
- Social: interacts with others (cooperation, communication, coordination, competition)

Features of agents

- Embodied: has sensors and effectors to read from and make changes to the environment

Features of agents

- Embodied: has sensors and effectors to read from and make changes to the environment
- Intelligent:

Features of agents

- Embodied: has sensors and effectors to read from and make changes to the environment
- Intelligent: ...whatever it means

Features of agents

- Embodied: has sensors and effectors to read from and make changes to the environment
- Intelligent: ...whatever it means
- Rational: always does the right thing

Is there any essential (and commonly accepted) feature of an agent?

Is there any essential (and commonly accepted) feature of an agent?

An agent acts.

Is there any essential (and commonly accepted) feature of an agent?

An agent acts.

Agents can be described mathematically by a function

$$act : \text{set of percept sequences} \mapsto \text{set of actions}$$

Is there any essential (and commonly accepted) feature of an agent?

An agent acts.

Agents can be described mathematically by a function

$$act : \text{set of percept sequences} \mapsto \text{set of actions}$$

Note that, in game theory, such a function is called a strategy.

Is there any essential (and commonly accepted) feature of an agent?

An agent acts.

Agents can be described mathematically by a function

$$act : \text{set of percept sequences} \mapsto \text{set of actions}$$

Note that, in game theory, such a function is called a strategy.

In planning, it is called a conditional plan.

1.2 Logic for Agents

Recall...

Multi-agent systems:

- A paradigm for thinking and talking about the world
- Makes us use specific vocabulary
- Makes us use specific conceptual structures
- Provides methodology for design and programming

Recall...

Multi-agent systems:

- A paradigm for thinking and talking about the world
- Makes us use specific vocabulary
- Makes us use specific conceptual structures
- Provides methodology for design and programming

This view of MAS comes close to the role of logic in both philosophy and computer science!

Logic:

- A paradigm for modeling and reasoning about the world in a precise manner

Logic:

- A paradigm for modeling and reasoning about the world in a precise manner
- Provides vocabulary and conceptual structures

Logic:

- A paradigm for modeling and reasoning about the world in a precise manner
- Provides vocabulary and conceptual structures
- Provides methodology for specification and verification

Logic:

- A paradigm for modeling and reasoning about the world in a precise manner
- Provides vocabulary and conceptual structures
- Provides methodology for specification and verification

Can be used for practical things (also in MAS):

Logic:

- A paradigm for modeling and reasoning about the world in a precise manner
- Provides vocabulary and conceptual structures
- Provides methodology for specification and verification

Can be used for practical things (also in MAS):

- automatic verification
- executable specifications
- planning as model checking

Logic and MAS can be a good match.

1.3 Modal Logic

Modal logic is an extension of classical logic by new connectives $\Box$ and $\Diamond$: necessity and possibility.

Modal logic is an extension of classical logic by new connectives $\Box$ and $\Diamond$: necessity and possibility.

- $\Box\varphi$ means that φ is necessarily true

Modal logic is an extension of classical logic by new connectives $\Box$ and $\Diamond$: necessity and possibility.

- $\Box\varphi$ means that φ is necessarily true
- $\Diamond\varphi$ means that φ is possibly true

Modal logic is an extension of classical logic by new connectives $\Box$ and $\Diamond$: necessity and possibility.

- $\Box\varphi$ means that φ is necessarily true
- $\Diamond\varphi$ means that φ is possibly true

Independently of the precise definition, the following holds:

$$\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$$

Definition 1.1 (Modal Logic with n modalities)

The language of modal logic with n modal operators $\Box_1, \ldots, \Box_n$ is the smallest set containing:

- atomic propositions $p, q, r, \ldots$;
- for formulae φ, it also contains $\neg\varphi, \Box_1\varphi, \ldots, \Box_n\varphi$;
- for formulae φ, ψ, it also contains $\varphi \wedge \psi$.

We treat $\vee, \rightarrow, \leftrightarrow, \Diamond$ as macros (defined as usual).

Definition 1.1 (Modal Logic with n modalities)

The language of modal logic with n modal operators $\Box_1, \ldots, \Box_n$ is the smallest set containing:

- atomic propositions $p, q, r, \ldots$;
- for formulae φ, it also contains $\neg\varphi, \Box_1\varphi, \ldots, \Box_n\varphi$;
- for formulae φ, ψ, it also contains $\varphi \wedge \psi$.

We treat $\vee, \rightarrow, \leftrightarrow, \Diamond$ as macros (defined as usual).

Note that the modal operators can be nested:

$$(\Box_1\Box_2\Diamond_1 p) \;\vee\; \Box_3\neg p$$

More precisely, necessity/possibility is interpreted as follows:

- p is necessary $\Leftrightarrow$ p is true in all possible scenarios
- p is possible $\Leftrightarrow$ p is true in at least one possible scenario

More precisely, necessity/possibility is interpreted as follows:

- p is necessary $\Leftrightarrow$ p is true in all possible scenarios
- p is possible $\Leftrightarrow$ p is true in at least one possible scenario

$\rightsquigarrow$ possible worlds semantics

Definition 1.2 (Kripke Structure)

A Kripke structure is a tuple $\langle \mathcal{W}, \mathcal{R} \rangle$, where $\mathcal{W}$ is a set of possible worlds, and $\mathcal{R}$ is a binary relation on worlds, called accessibility relation.

Definition 1.2 (Kripke Structure)

A Kripke structure is a tuple $\langle \mathcal{W}, \mathcal{R} \rangle$, where $\mathcal{W}$ is a set of possible worlds, and $\mathcal{R}$ is a binary relation on worlds, called accessibility relation.

Definition 1.3 (Kripke model)

A possible worlds model $\mathcal{M} = \langle \mathcal{S}, \pi \rangle$ consists of a Kripke structure $\mathcal{S}$, and a valuation of propositions $\pi : \mathcal{W} \rightarrow \mathcal{P}(\{p, q, r, \ldots\})$.

Remarks:

- $\mathcal{R}$ indicates which worlds are relevant for each other; $w_1 \mathcal{R} w_2$ can be read as "world w_2 is relevant for (reachable from) world w_1"

Remarks:

- $\mathcal{R}$ indicates which worlds are relevant for each other; $w_1 \mathcal{R} w_2$ can be read as "world w_2 is relevant for (reachable from) world w_1"
- $\mathcal{R}$ can be any binary relation from $\mathcal{W} \times \mathcal{W}$; we do not require any specific properties (yet).

Remarks:

- $\mathcal{R}$ indicates which worlds are relevant for each other; $w_1 \mathcal{R} w_2$ can be read as "world w_2 is relevant for (reachable from) world w_1"
- $\mathcal{R}$ can be any binary relation from $\mathcal{W} \times \mathcal{W}$; we do not require any specific properties (yet).
- It is natural to see the worlds from $\mathcal{W}$ as classical propositional models, i.e. valuations of propositions $\pi(w) \subseteq \{p, q, r, \ldots\}$.

Definition 1.4 (Semantics of modal logic)

The truth of formulae is relative to a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \pi \rangle$, and a world $w \in \mathcal{W}$. It can be defined through the following clauses:

Definition 1.4 (Semantics of modal logic)

The truth of formulae is relative to a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \pi \rangle$, and a world $w \in \mathcal{W}$. It can be defined through the following clauses:

- $\mathcal{M}, w \models p$ iff $p \in \pi(w)$;

Definition 1.4 (Semantics of modal logic)

The truth of formulae is relative to a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \pi \rangle$, and a world $w \in \mathcal{W}$. It can be defined through the following clauses:

- $\mathcal{M}, w \models p$ iff $p \in \pi(w)$;
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$;

Definition 1.4 (Semantics of modal logic)

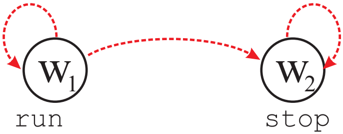
The truth of formulae is relative to a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \pi \rangle$, and a world $w \in \mathcal{W}$. It can be defined through the following clauses:

- $\mathcal{M}, w \models p$ iff $p \in \pi(w)$;
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$;
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$;

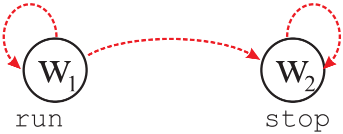
Definition 1.4 (Semantics of modal logic)

The truth of formulae is relative to a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \pi \rangle$, and a world $w \in \mathcal{W}$. It can be defined through the following clauses:

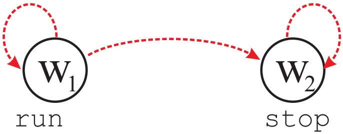
- $\mathcal{M}, w \models p$ iff $p \in \pi(w)$;
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$;
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$;
- $\mathcal{M}, w \models \Box\varphi$ iff, for every $w' \in \mathcal{W}$ such that $w\mathcal{R}w'$, we have $\mathcal{M}, w' \models \varphi$.



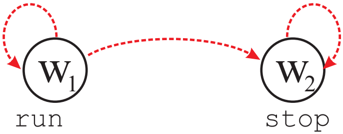
w_1
run
w_2
stop



$$\mathsf{run} \longrightarrow \Diamond\mathsf{stop}$$



$$run \longrightarrow \Diamond stop$$
$$stop \longrightarrow \Box stop$$



$$\text{run} \longrightarrow \Diamond\text{stop}$$

$$\text{stop} \longrightarrow \Box\text{stop}$$

$$\text{run} \longrightarrow \Diamond\Box\text{stop}$$

Note:

- Most modal logics can be translated to classical logic

Note:

- Most modal logics can be translated to classical logic
- . . . but the result looks horribly ugly

Note:

- Most modal logics can be translated to classical logic
- . . . but the result looks horribly ugly,
- . . . and in most cases it is much harder to automatize anything

1.4 Axioms for Modal Logics

As in classical logic, one can ask about a complete axiom system. Is there a calculus that allows to derive all sentences that are true in all Kripke models?

As in classical logic, one can ask about a complete axiom system. Is there a calculus that allows to derive all sentences that are true in all Kripke models?

Definition 1.5 (System K)

System **K** is an extension of the propositional calculus by the axiom

$$\mathbf{K} \quad (\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \rightarrow \Box\psi$$

and the inference rule

$$\textbf{(Necessitation)} \quad \frac{\varphi}{\Box\varphi}.$$

Theorem 1.6 (Soundness/completeness of system K)

*System **K** is sound and complete with respect to the class of all Kripke models.*

Theorem 1.6 (Soundness/completeness of system K)

*System **K** is sound and complete with respect to the class of all Kripke models.*

Note: with n modalities, the calculus is called $\mathbf{K}_n$, and the theorem extends in a straightforward way.

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

$$\mathbf{K} \quad (\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \rightarrow \Box\psi$$

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

K	$(\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \rightarrow \Box\psi$
D	$\neg\Box(\varphi \wedge \neg\varphi)$

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

K	$(\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \rightarrow \Box\psi$
D	$\neg\Box(\varphi \wedge \neg\varphi)$
T	$\Box\varphi \rightarrow \varphi$

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

K	$(\Box\varphi \wedge \Box(\varphi \to \psi)) \to \Box\psi$
D	$\neg\Box(\varphi \wedge \neg\varphi)$
T	$\Box\varphi \to \varphi$
4	$\Box\varphi \to \Box\Box\varphi$

Definition 1.7 (Extending K with axioms D, T, 4, 5)

System **K** is often extended by (a subset of) the following axioms (called as below for historical reasons):

K	$(\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \rightarrow \Box\psi$
D	$\neg\Box(\varphi \wedge \neg\varphi)$
T	$\Box\varphi \rightarrow \varphi$
4	$\Box\varphi \rightarrow \Box\Box\varphi$
5	$\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$

Best known extensions of system **K**:

- **S5** = **KDT45**: the standard logic of knowledge
- **KD45**: the standard logic of beliefs

Theorem 1.8 (Sound/complete subsystems of KDT45)

Let $\mathbf{X}$ *be any subset of* $\{\mathbf{D}, \mathbf{T}, \mathbf{4}, \mathbf{5}\}$ *and let* $\mathcal{X}$ *be any subset of* {*serial*, *reflexive*, *transitive*, *euclidean*} *corresponding to* $\mathbf{X}$.
Then $\mathbf{K} \cup \mathbf{X}$ *is sound and complete with respect to Kripke models the accessibility relation of which satisfies* $\mathcal{X}$.

Theorem 1.8 (Sound/complete subsystems of KDT45)

Let $\mathbf{X}$ *be any subset of* $\{\mathbf{D}, \mathbf{T}, \mathbf{4}, \mathbf{5}\}$ *and let* $\mathcal{X}$ *be any subset of* $\{serial, reflexive, transitive, euclidean\}$ *corresponding to* $\mathbf{X}$.
Then $\mathbf{K} \cup \mathbf{X}$ *is sound and complete with respect to Kripke models the accessibility relation of which satisfies* $\mathcal{X}$.

Corollary 1.9

System ***S5*** *is sound and complete with respect to Kripke models with equivalence accessibility relations.*

Exercise

Show that

1. Axiom **D** follows from **KT45**.
2. **KD45** is not equivalent to **K45**: axiom **D** does not follow from **K45**.

1.5 Methodology

When is a formula true?

When is a formula true?

- φ can be true in $\mathcal{M}$ and q $(\mathcal{M}, q \models \varphi)$

When is a formula true?

- φ can be true in $\mathcal{M}$ and q ($\mathcal{M}, q \models \varphi$)
- φ can be valid in $\mathcal{M}$ ($\mathcal{M}, q \models \varphi$ for all q)

When is a formula true?

- φ can be true in $\mathcal{M}$ and q ($\mathcal{M}, q \models \varphi$)
- φ can be valid in $\mathcal{M}$ ($\mathcal{M}, q \models \varphi$ for all q)
- φ can be valid ($\mathcal{M}, q \models \varphi$ for all $\mathcal{M}, q$)

When is a formula true?

- φ can be true in $\mathcal{M}$ and q ($\mathcal{M}, q \models \varphi$)
- φ can be valid in $\mathcal{M}$ ($\mathcal{M}, q \models \varphi$ for all q)
- φ can be valid ($\mathcal{M}, q \models \varphi$ for all $\mathcal{M}, q$)
- φ can be satisfiable ($\mathcal{M}, q \models \varphi$ for some $\mathcal{M}, q$)

When is a formula true?

- φ can be true in $\mathcal{M}$ and q ($\mathcal{M}, q \models \varphi$)
- φ can be valid in $\mathcal{M}$ ($\mathcal{M}, q \models \varphi$ for all q)
- φ can be valid ($\mathcal{M}, q \models \varphi$ for all $\mathcal{M}, q$)
- φ can be satisfiable ($\mathcal{M}, q \models \varphi$ for some $\mathcal{M}, q$)
- φ can be a theorem (it can be derived from the axioms via inference rules)

Methodology $\rightleftharpoons$ Problems...

Methodology $\rightleftharpoons$ Problems... in a positive sense

Methodology $\rightleftharpoons$ Problems... in a positive sense

$=$ questions that can be asked with logic:

Methodology $\rightleftharpoons$ Problems... in a positive sense
= questions that can be asked with logic:

- model checking (local): "given $\mathcal{M}$, q, and φ, is φ true in $\mathcal{M}, q$?"

Methodology $\rightleftharpoons$ Problems... in a positive sense
= questions that can be asked with logic:

- model checking (local): "given $\mathcal{M}$, q, and φ, is φ true in $\mathcal{M}, q$?"
- model checking (global): "given $\mathcal{M}$ and φ, what is the set of states in which φ is true?"

Methodology $\rightleftharpoons$ Problems... in a positive sense
= questions that can be asked with logic:

- model checking (local): "given $\mathcal{M}$, q, and φ, is φ true in $\mathcal{M}, q$?"
- model checking (global): "given $\mathcal{M}$ and φ, what is the set of states in which φ is true?"
- satisfiability: "given φ, is φ true in at least one model and state?"

Methodology $\rightleftharpoons$ Problems... in a positive sense
= questions that can be asked with logic:

- model checking (local): "given $\mathcal{M}$, q, and φ, is φ true in $\mathcal{M}, q$?"
- model checking (global): "given $\mathcal{M}$ and φ, what is the set of states in which φ is true?"
- satisfiability: "given φ, is φ true in at least one model and state?"
- validity: "given φ, is φ true in all models and their states?"

Methodology $\rightleftharpoons$ Problems... in a positive sense
= questions that can be asked with logic:

- model checking (local): "given $\mathcal{M}$, q, and φ, is φ true in $\mathcal{M}, q$?"
- model checking (global): "given $\mathcal{M}$ and φ, what is the set of states in which φ is true?"
- satisfiability: "given φ, is φ true in at least one model and state?"
- validity: "given φ, is φ true in all models and their states?"
- theorem proving: "given φ, is it possible to prove (derive) φ?"

1.6 Logic for Agents ctd.

Modal logic is a generic framework.

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,
- actions $\rightsquigarrow$ dynamic logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,
- actions $\rightsquigarrow$ dynamic logic,
- time $\rightsquigarrow$ temporal logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,
- actions $\rightsquigarrow$ dynamic logic,
- time $\rightsquigarrow$ temporal logic,
- ability $\rightsquigarrow$ strategic logic,

Modal logic is a generic framework.

Various modal logics:

- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,
- actions $\rightsquigarrow$ dynamic logic,
- time $\rightsquigarrow$ temporal logic,
- ability $\rightsquigarrow$ strategic logic,
- and combinations of the above

Modal logic seems very well suited for reasoning about various dimensions of multi-agent systems!

Modal logic seems very well suited for reasoning about various dimensions of multi-agent systems!

Modal logic and MAS *are* a good match!