A4M33MAS - Multiagent Systems
Agents and their behavior modeling by means of formal logic

Michal Pechoucek & Michal Jakob

Department of Computer Science
Czech Technical University in Prague

In parts based an selected graphics taken from Valentin Goranko and Wojtek Jamroga: Modal Logics for Multi-Agent Systems, 8th European Summer School in Logic Language and Information

Multi-agent systems & Logic

- Multi-agent systems
  - Complex decentralized systems whose behaviour is given by interaction among autonomous, rational entities. We study MAS so that we understand behaviour of such systems and can design such software systems.

- Logic
  - Provides a paradigm for modeling and reasoning about the complex world in a precise and exact manner
  - Provides methodology for specification and verification of complex programs

- Can be used for practical things (also in MAS):
  - automatic verification of multi-agent systems
  - and/or executable specifications of multi-agent systems

Best logic for MAS?

Modal logic

01

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:

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

Modal logic syntax

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:

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

Modal logic syntax

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

Modal logic 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 1.3 (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 $\pi : \mathcal{W} \rightarrow \mathcal{P}(\{p, q, r, \ldots\})$.

Modal logic semantics

0I

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).

Modal logic semantics

OI

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 $\pi(w) \subseteq \{p, q, r, \ldots\}$.

Modal logic semantics

OI

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 $\mathcal{M}, w' \models \varphi$.

Modal logic example

Modal logic example

$$run \rightarrow \Diamond stop$$

Modal logic example

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

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

Modal logic example

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

Modal logic

- 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

Axioms in Modal logic

Definition 1.5 (System K)

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

Distribution axiom

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

and the inference rule

Generalization axiom $$\frac{\varphi}{\Box\varphi}.$$

Axioms in Modal logic

Theorem 1.6 (Soundness/completeness of system K)

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

Axioms in Modal logic

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):

- T: $\Box\varphi \rightarrow \varphi$
- D: $\Box\varphi \rightarrow \Diamond\varphi$
- 4: $\Box\varphi \rightarrow \Box\Box\varphi$
- B: $\varphi \rightarrow \Box\Diamond\varphi$
- 5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Proofs

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ⊚

$$\text{T: } \Box\varphi \rightarrow \varphi$$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ⊚

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R \Rightarrow \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ⊚

$$\text{D: } \Box\varphi \rightarrow \Diamond\varphi$$

Proofs

01

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi \; \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

D: $\Box\varphi \rightarrow \Diamond\varphi$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w,w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w,w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w,w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w,w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

$$\text{D: } \Box\varphi \rightarrow \Diamond\varphi$$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

4: $\Box\varphi \rightarrow \Box\Box\varphi$

Proofs

01

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

$$4: \Box\varphi \rightarrow \Box\Box\varphi$$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ⊚

D: $(\mathcal{M}_1 \models_w \varphi \; \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ⊚

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi \; \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ⊚

$$4: \; \Box\varphi \rightarrow \Box\Box\varphi$$

Proofs

01

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

$$\text{B: } \varphi \rightarrow \Box\Diamond\varphi$$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi \; \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi \; \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

B: $\varphi \rightarrow \Box\Diamond\varphi$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

$$\text{B: } \varphi \rightarrow \Box\Diamond\varphi$$

Proofs

01

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ⊚

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ⊚

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ⊚

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ⊚

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w' (w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ⊚

$$5:\ \Diamond\varphi \rightarrow \Box\Diamond\varphi$$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ⊚

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ⊚

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ⊚

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ⊚

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w'(w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ⊚

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w'(w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ◎

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w'(w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ◎

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Proofs

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w'(w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ◎

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Proofs

01

T: because $\models \varphi \Rightarrow \Box\varphi$ and due reflexivity $\forall w : (w, w) \in R$ ◎

D: $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ and due to seriality $(\mathcal{M}_1 \models_w (\exists w' : (w, w') \in R))$ we can say that $\mathcal{M}_1 \models_w \exists w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \varphi)$ ◎

4: provided that there is transitive relation on R we may say that $(\mathcal{M}_1 \models_w \varphi\ \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} (\forall w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi))$ ◎

B: provided that there is symetric relation on R we say that $\mathcal{M}_1 \models_w \varphi \Rightarrow \forall w' : (w, w') \in R : \mathcal{M}_1 \models_{w'} \exists w'' : (w', w'') \in R : \mathcal{M}_1 \models_{w''} \varphi$ if $(\forall w, w', (w, w') \in R \Rightarrow (w', w) \in R)$ then $w = w''$ and $\mathcal{M}_1 \models_w \varphi$ ◎

5: $(\mathcal{M}_1 \models_w \exists w' : (w, w') \in R \models_{w'} \varphi) \Rightarrow (\mathcal{M}_1 \models_w \forall w'' : (w, w'') \in R : \mathcal{M}_1 \models_{w''} \exists w' (w'', w') \in R : \mathcal{M}_1 \models_{w'} \varphi)$ due to euclidean property if $(w, w') \in R \wedge (w, w'') \in R$ then $(w', w'') \in R$ ◎

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Axioms in Modal logic

01

- T: $\Box\varphi \rightarrow \varphi$ due to reflexivity
- D: $\Box\varphi \rightarrow \Diamond\varphi$ due to seriality
- 4: $\Box\varphi \rightarrow \Box\Box\varphi$ due to transitivity
- B: $\varphi \rightarrow \Box\Diamond\varphi$ due to symetricity
- 5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$ due to euclidean property

Model of Belief & Knowledge

Model of Belief & Knowledge

- Once we are implementing an intelligent agent what do we want the program to implement e.g. its beliefs:
  - to satisfy the K axioms
  - *an agent knows what it does know*: positive introspection axiom (4 axiom).
  - *an agent knows what it does not know*: negative introspection axiom (5 axiom).
  - *it beliefs are not contradictory*: if it knows something it means it does not allow the negation of its being true (D axiom).

T: $\Box\varphi \rightarrow \varphi$

D: $\Box\varphi \rightarrow \Diamond\varphi$

4: $\Box\varphi \rightarrow \Box\Box\varphi$

B: $\varphi \rightarrow \Box\Diamond\varphi$

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Model of Belief & Knowledge

- Once we are implementing an intelligent agent what do we want the program to implement e.g. its beliefs:
  - to satisfy the K axioms
  - *an agent knows what it does know*: positive introspection axiom (4 axiom).
  - *an agent knows what it does not know*: negative introspection axiom (5 axiom).
  - *it beliefs are not contradictory*: if it knows something it means it does not allow the negation of its being true (D axiom).

- Belief is surely a KD45 system -- modal logic system where the B relation is serial, transitive and euclidean.

T: $\Box\varphi \rightarrow \varphi$

D: $\Box\varphi \rightarrow \Diamond\varphi$

4: $\Box\varphi \rightarrow \Box\Box\varphi$

B: $\varphi \rightarrow \Box\Diamond\varphi$

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Model of Belief & Knowledge

01

- Once we are implementing an intelligent agent what do we want the program to implement e.g. its beliefs:
  - to satisfy the K axioms
  - *an agent knows what it does know*: positive introspection axiom (4 axiom).
  - *an agent knows what it does not know*: negative introspection axiom (5 axiom).
  - *it beliefs are not contradictory*: if it knows something it means it does not allow the negation of its being true (D axiom).

- Belief is surely a KD45 system -- modal logic system where the B relation is serial, transitive and euclidean.

- Knowledge is more difficult – it needs to be also true – this why the knowledge accessibility relation needs to be also reflexive.

T: $\Box\varphi \rightarrow \varphi$

D: $\Box\varphi \rightarrow \Diamond\varphi$

4: $\Box\varphi \rightarrow \Box\Box\varphi$

B: $\varphi \rightarrow \Box\Diamond\varphi$

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Model of Belief & Knowledge

- Once we are implementing an intelligent agent what do we want the program to implement e.g. its beliefs:
  - to satisfy the K axioms
  - *an agent knows what it does know*: positive introspection axiom (4 axiom).
  - *an agent knows what it does not know*: negative introspection axiom (5 axiom).
  - *it beliefs are not contradictory*: if it knows something it means it does not allow the negation of its being true (D axiom).

- Belief is surely a KD45 system -- modal logic system where the B relation is serial, transitive and euclidean.

- Knowledge is more difficult – it needs to be also true – this why the knowledge accessibility relation needs to be also reflexive.
- Therefore knowledge is a KTD45 system.

T: $\Box\varphi \rightarrow \varphi$

D: $\Box\varphi \rightarrow \Diamond\varphi$

4: $\Box\varphi \rightarrow \Box\Box\varphi$

B: $\varphi \rightarrow \Box\Diamond\varphi$

5: $\Diamond\varphi \rightarrow \Box\Diamond\varphi$

Automated reasoning in Logic

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

Automated reasoning in Logic

- 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)

Automated reasoning in Logic

- 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 $\mathcal{M}, q$)

Automated reasoning in Logic

- 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 $\mathcal{M}, q$)

Automated reasoning in Logic

- 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)

Automated reasoning in 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?"

Model checking is a technique for automatically verifying correctness properties of finite-state systems. Given a model of a system, exhaustively and automatically check whether this model meets a given specification (such as the absence of deadlocks and similar critical states that can cause the system to crash).

Automated reasoning in 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?"

Various Modal Logics

Modal logic is a generic framework.

Various modal logics:

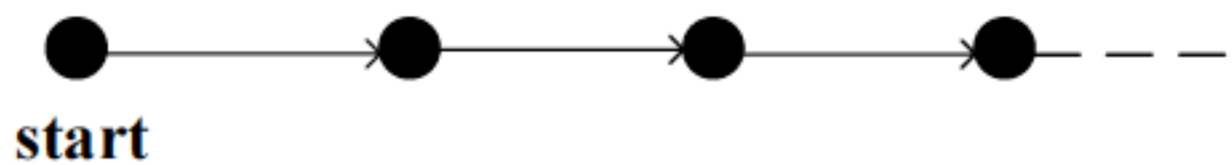
- knowledge $\rightsquigarrow$ epistemic logic,
- beliefs $\rightsquigarrow$ doxastic logic,
- obligations $\rightsquigarrow$ deontic logic,
- actions $\rightsquigarrow$ dynamic logic,
- time $\rightsquigarrow$ temporal logic,
- ability $\rightsquigarrow$ strategic logic,
- and combinations of the above

Model of Time

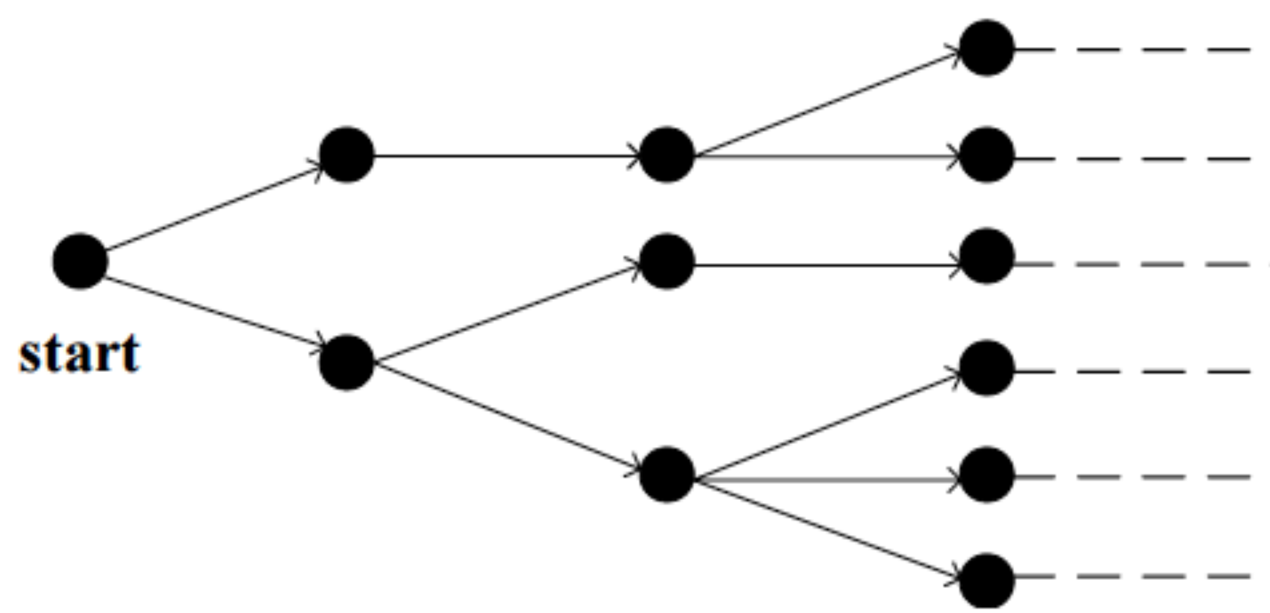
Model of Time

- Modeling time as an instance of modal logic where the accessibility relation represents the relationship between the past, current and future time moments.

- Time:
  - linear



  - branching



Typical Temporal Operators

$\mathcal{X}\varphi$	φ is true in the next moment in time
$\mathcal{G}\varphi$	φ is true in all future moments
$\mathcal{F}\varphi$	φ is true in some future moment
$\varphi\mathcal{U}\psi$	φ is true until the moment when ψ becomes true

$$\mathcal{G}((\neg \mathsf{passport} \vee \neg \mathsf{ticket}) \rightarrow \mathcal{X}\neg \mathsf{board\_flight})$$

$$\mathsf{send}(\mathsf{msg}, \mathsf{rcvr}) \rightarrow \mathcal{F}\mathsf{receive}(\mathsf{msg}, \mathsf{rcvr})$$

Safety Property

- *something bad will not happen*
- *something good will always hold*

Safety Property

– *something bad will not happen*
– *something good will always hold*

- Typical examples:

$$\mathcal{G}\neg\mathsf{bankrupt}$$

Safety Property

– *something bad will not happen*
– *something good will always hold*

- Typical examples

$\mathcal{G}\neg\text{bankrupt}$

$\mathcal{G}(\text{fuelOK} \vee \mathcal{X}\text{fuelOK})$

and so on . . .

Safety Property

- *something bad will not happen*
- *something good will always hold*

- Typical examples

  $\mathcal{G}\neg\text{bankrupt}$

  $\mathcal{G}(\text{fuelOK} \vee \mathcal{X}\text{fuelOK})$

  and so on ...

Usually: $\mathcal{G}\neg\ldots$

Liveness Property

– *something good will happen*

Liveness Property

– *something good will happen*

- Typical examples

$\mathcal{F}\text{rich}$

Liveness Property

01

– *something good will happen*

- Typical examples

$\mathcal{F}\mathsf{rich}$

$\mathsf{rocketLondon} \rightarrow \mathcal{F}\mathsf{rocketParis}$

and so on . . .

Liveness Property

– *something good will happen*

- Typical examples

$$\mathcal{F}\mathsf{rich}$$

$$\mathsf{rocketLondon} \rightarrow \mathcal{F}\mathsf{rocketParis}$$

and so on ...

Usually: $\mathcal{F}$....

Fairness Property

- *Useful when scheduling processes, responding to messages, etc.*
- *Good for specifying interaction properties of the environment*

- Typical examples:

$$\mathcal{G}(\text{rocketLondon} \rightarrow \mathcal{F}\text{rocketParis})$$

- Strong Fairness:
  if something is attempted/requested, then it will be successful

- Typical examples:

$$\mathcal{G}(\text{attempt} \rightarrow \mathcal{F}\text{success})$$

$$\mathcal{G}\mathcal{F}\text{attempt} \rightarrow \mathcal{G}\mathcal{F}\text{success}$$

Linear Temporal Logic - LTL

- Reasoning about a particular computation of a system where time is linear - just one possible future path is included.

Definition 3.4 (Models of LTL)

A model of LTL is a sequence of time moments. We call such models paths, and denote them by λ.

Evaluation of atomic propositions at particular time moments is also needed.

Notation:

- $\lambda[i]$: ith time moment
- $\lambda[i \dots j]$: all time moments between i and j
- $\lambda[i \dots \infty]$: all timepoints from i on

Linear Temporal Logic - LTL

Definition 3.5 (Semantics of LTL)

$\lambda \models p$ iff p is true at moment $\lambda[0]$;

$\lambda \models \mathcal{X}\varphi$ iff $\lambda[1..\infty] \models \varphi$;

$\lambda \models \mathcal{F}\varphi$ iff $\lambda[i..\infty] \models \varphi$ for some $i \geq 0$;

$\lambda \models \mathcal{G}\varphi$ iff $\lambda[i..\infty] \models \varphi$ for all $i \geq 0$;

$\lambda \models \varphi\mathcal{U}\psi$ iff $\lambda[i..\infty] \models \psi$ for some $i \geq 0$, and $\lambda[j..\infty] \models \varphi$ for all $0 \leq j \leq i$.

Linear Temporal Logic - LTL

Definition 3.5 (Semantics of LTL)

$\lambda \models p$	iff p is true at moment $\lambda[0]$;
$\lambda \models \mathcal{X}\varphi$	iff $\lambda[1..\infty] \models \varphi$;
$\lambda \models \mathcal{F}\varphi$	iff $\lambda[i..\infty] \models \varphi$ for some $i \geq 0$;
$\lambda \models \mathcal{G}\varphi$	iff $\lambda[i..\infty] \models \varphi$ for all $i \geq 0$;
$\lambda \models \varphi\mathcal{U}\psi$	iff $\lambda[i..\infty] \models \psi$ for some $i \geq 0$, and $\lambda[j..\infty] \models \varphi$ for all $0 \leq j \leq i$.

Note that:

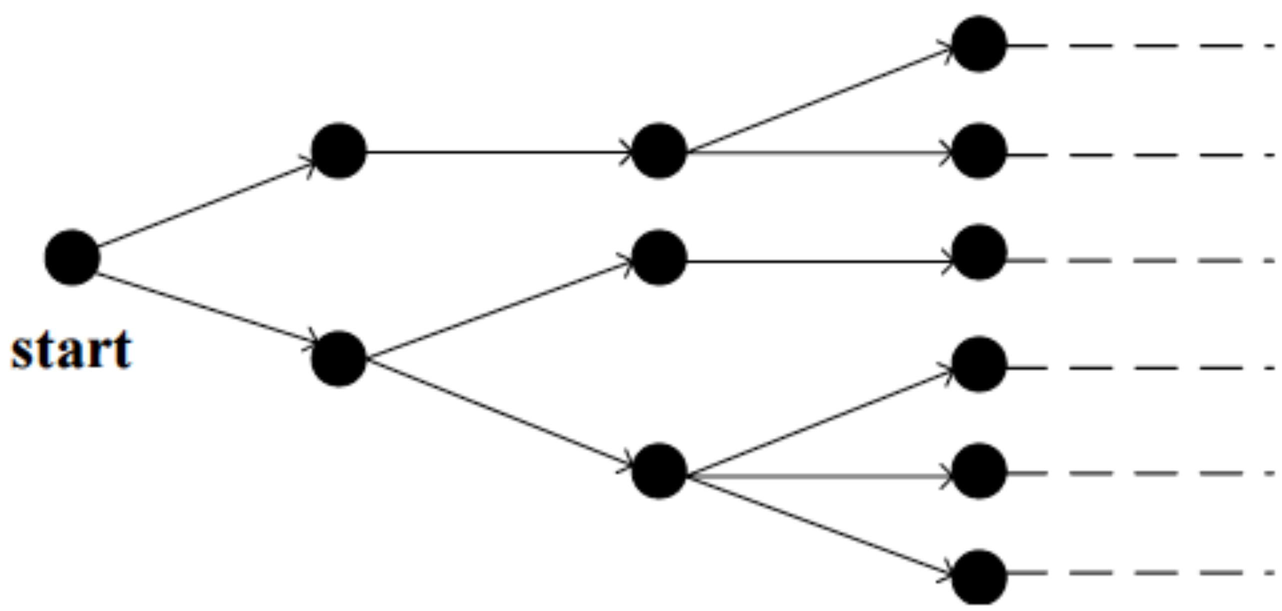
$$\mathcal{G}\varphi \equiv \neg\mathcal{F}\neg\varphi$$
$$\mathcal{F}\varphi \equiv \neg\mathcal{G}\neg\varphi$$
$$\mathcal{F}\varphi \equiv \top\,\mathcal{U}\varphi$$

Computational Tree Logic - CTL

- Reasoning about possible computations of a system. Time is branching – we want all alternative paths included.

Path quantifiers: **A** (for all paths), **E** (there is a path);

Temporal operators: $\mathcal{X}$ (nexttime), $\mathcal{F}$ (sometime), $\mathcal{G}$ (always) and $\mathcal{U}$ (until);



Computational Tree Logic - CTL

- Reasoning about possible computations of a system. Time is branching – we want all alternative paths included.

  Path quantifiers: **A** (for all paths), **E** (there is a path);
  Temporal operators: $\mathcal{X}$ (nexttime), $\mathcal{F}$ (sometime), $\mathcal{G}$ (always) and $\mathcal{U}$ (until);

- Vanilla CTL: every temporal operator must be immediately preceded by exactly one path quantifier
- CTL*: no syntactic restrictions
- Reasoning in Vanilla CTL can be automatized.

Computational Tree Logic - CTL

Definition 3.8 (Semantics of CTL*: state formulae)

$M, q \models \mathbf{E}\varphi$ iff there is a path λ, starting from q, such that $M, \lambda \models \varphi$;

$M, q \models \mathbf{A}\varphi$ iff for all paths λ, starting from q, we have $M, \lambda \models \varphi$.

Computational Tree Logic - CTL

0I

Definition 3.8 (Semantics of CTL*: state formulae)

$M, q \models \mathbf{E}\varphi$ iff there is a path λ, starting from q, such that $M, \lambda \models \varphi$;

$M, q \models \mathbf{A}\varphi$ iff for all paths λ, starting from q, we have $M, \lambda \models \varphi$.

Definition 3.9 (Semantics of CTL*: path formulae)

Exactly like for LTL!

Computational Tree Logic - CTL

Definition 3.8 (Semantics of CTL*: state formulae)

$M, q \models \mathbf{E}\varphi$ iff there is a path λ, starting from q, such that $M, \lambda \models \varphi$;

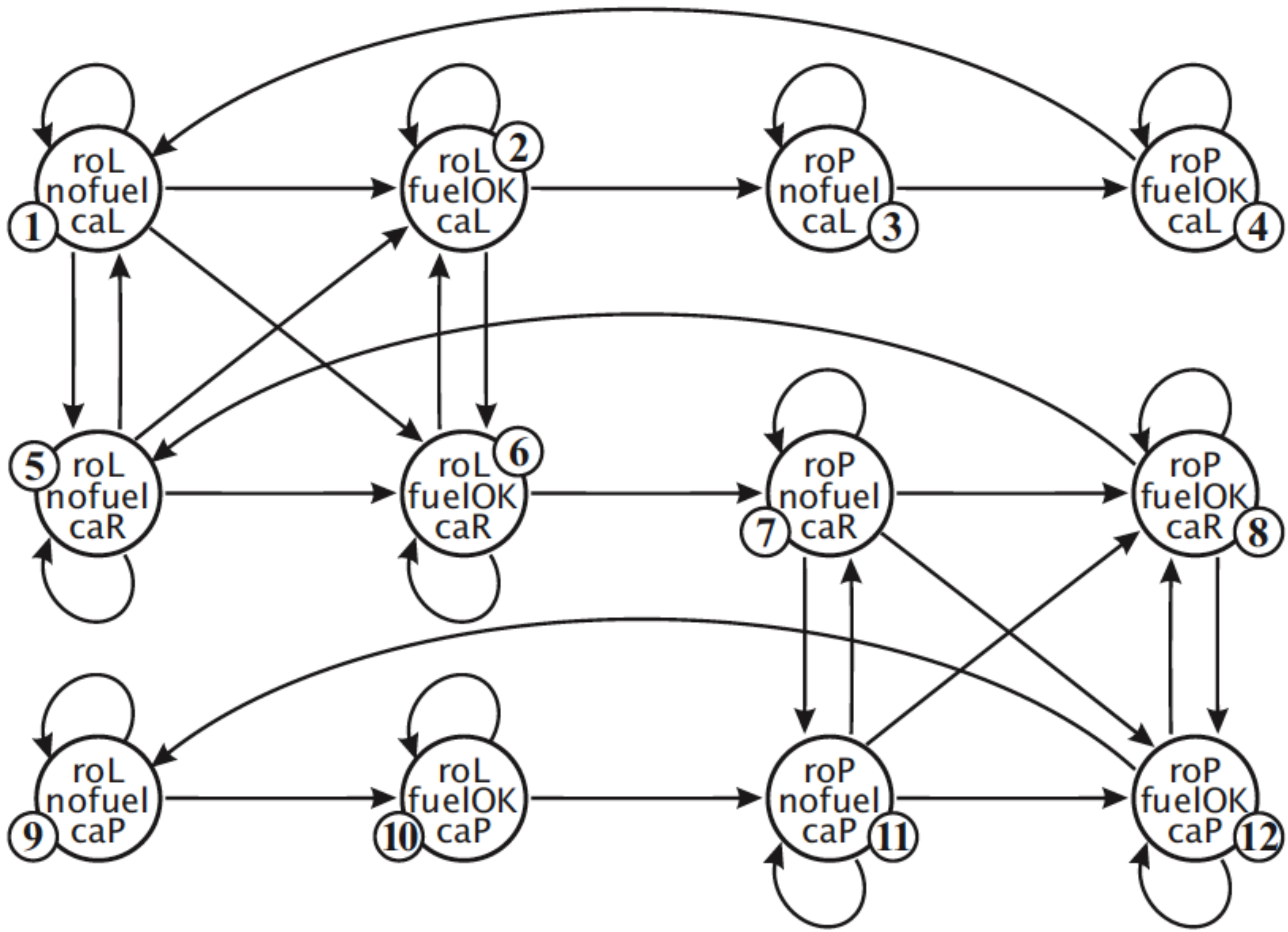
$M, q \models \mathbf{A}\varphi$ iff for all paths λ, starting from q, we have $M, \lambda \models \varphi$.

Definition 3.9 (Semantics of CTL*: path formulae)

$M, \lambda \models \mathcal{X}\varphi$ iff $M, \lambda[1 ... \infty] \models \varphi$;

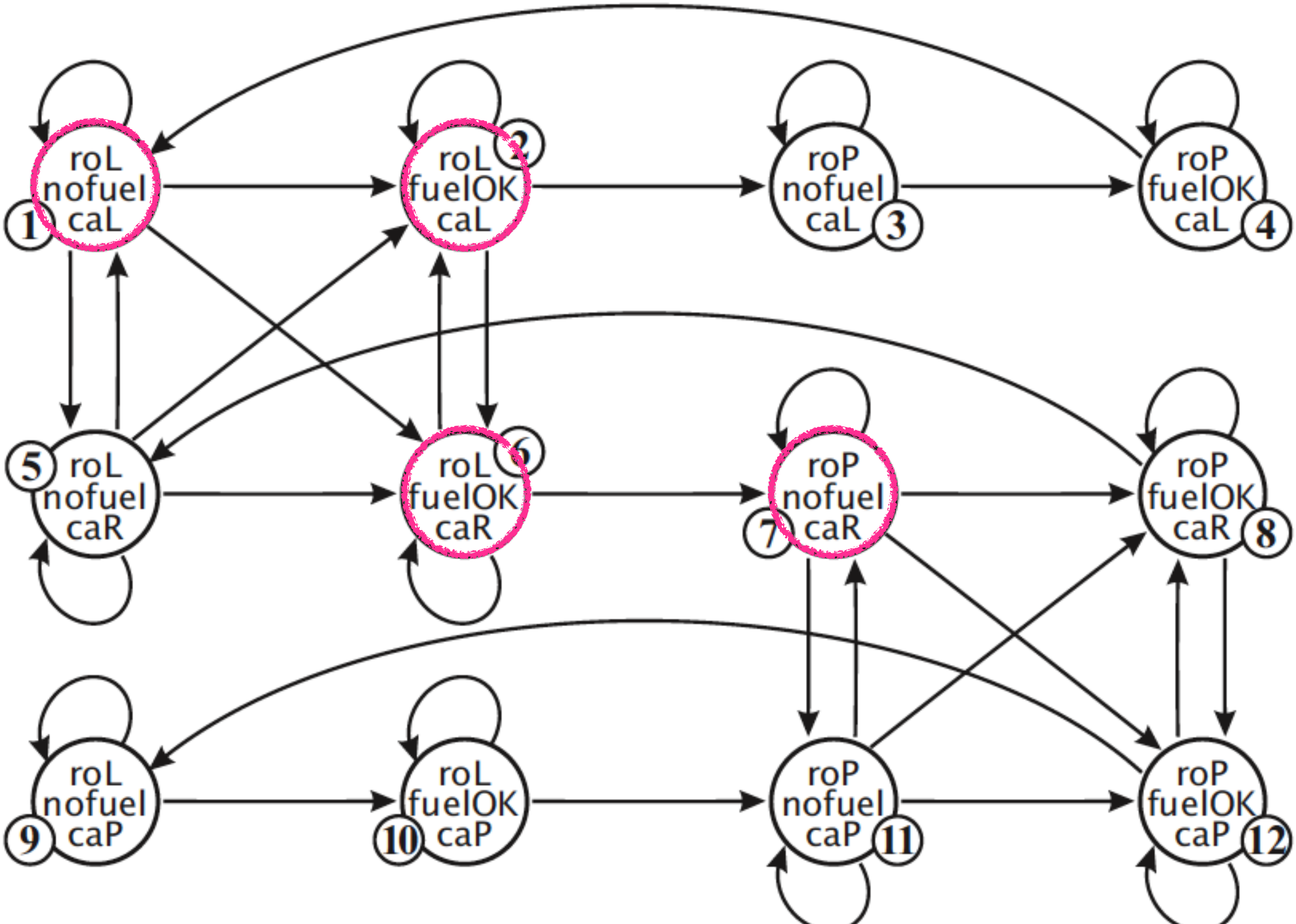
$M, \lambda \models \varphi \mathcal{U} \psi$ iff $M, \lambda[i ... \infty] \models \psi$ for some $i \geq 0$, and $M, \lambda[j ... \infty] \models \varphi$ for all $0 \leq j \leq i$.

Example

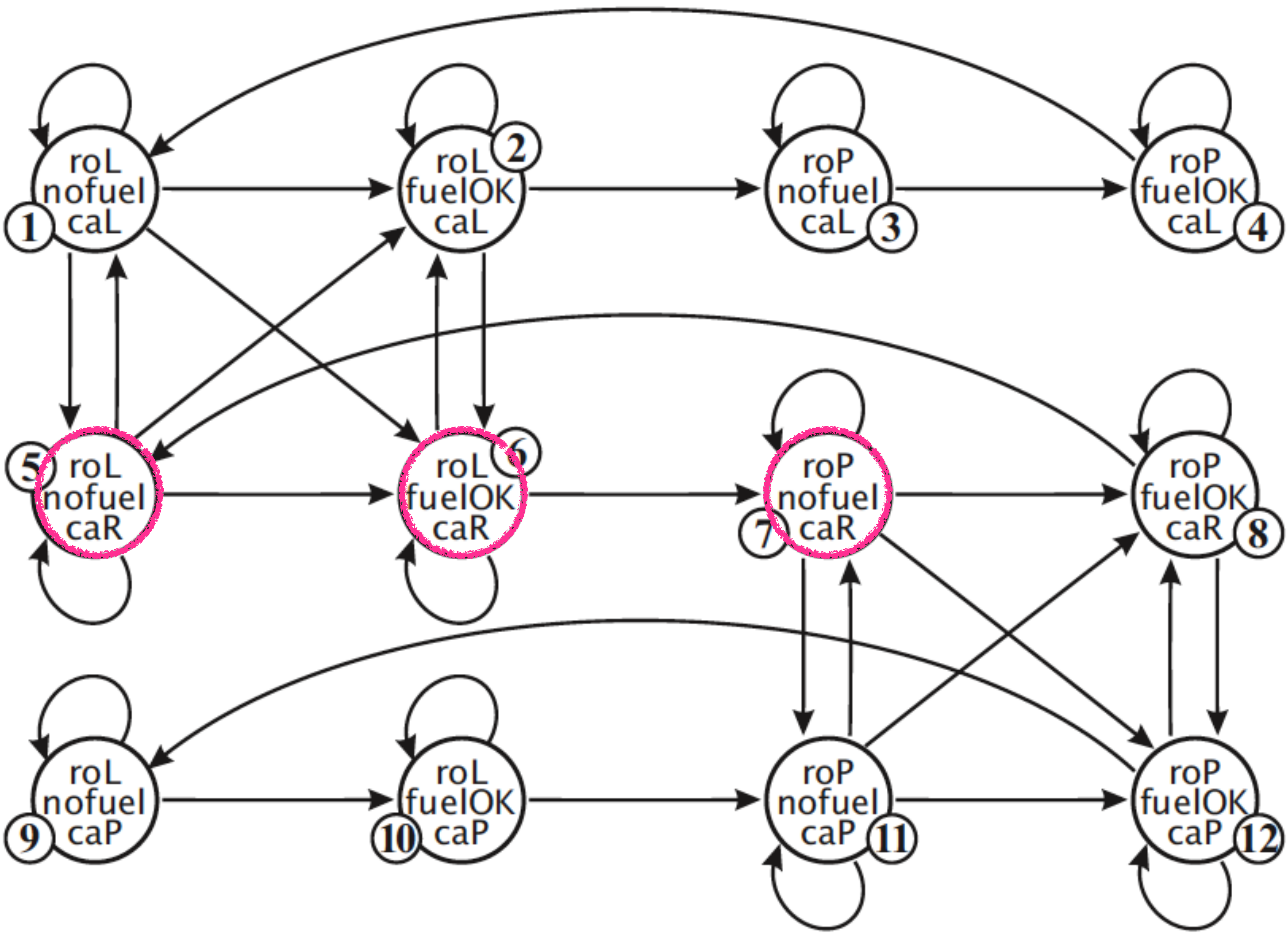


roL nofuel caL 1
roL fuelOK caL 2
roP nofuel caL 3
roP fuelOK caL 4
roL nofuel caR 5
roL fuelOK caR 6
roP nofuel caR 7
roP fuelOK caR 8
roL nofuel caP 9
roL fuelOK caP 10
roP nofuel caP 11
roP fuelOK caP 12

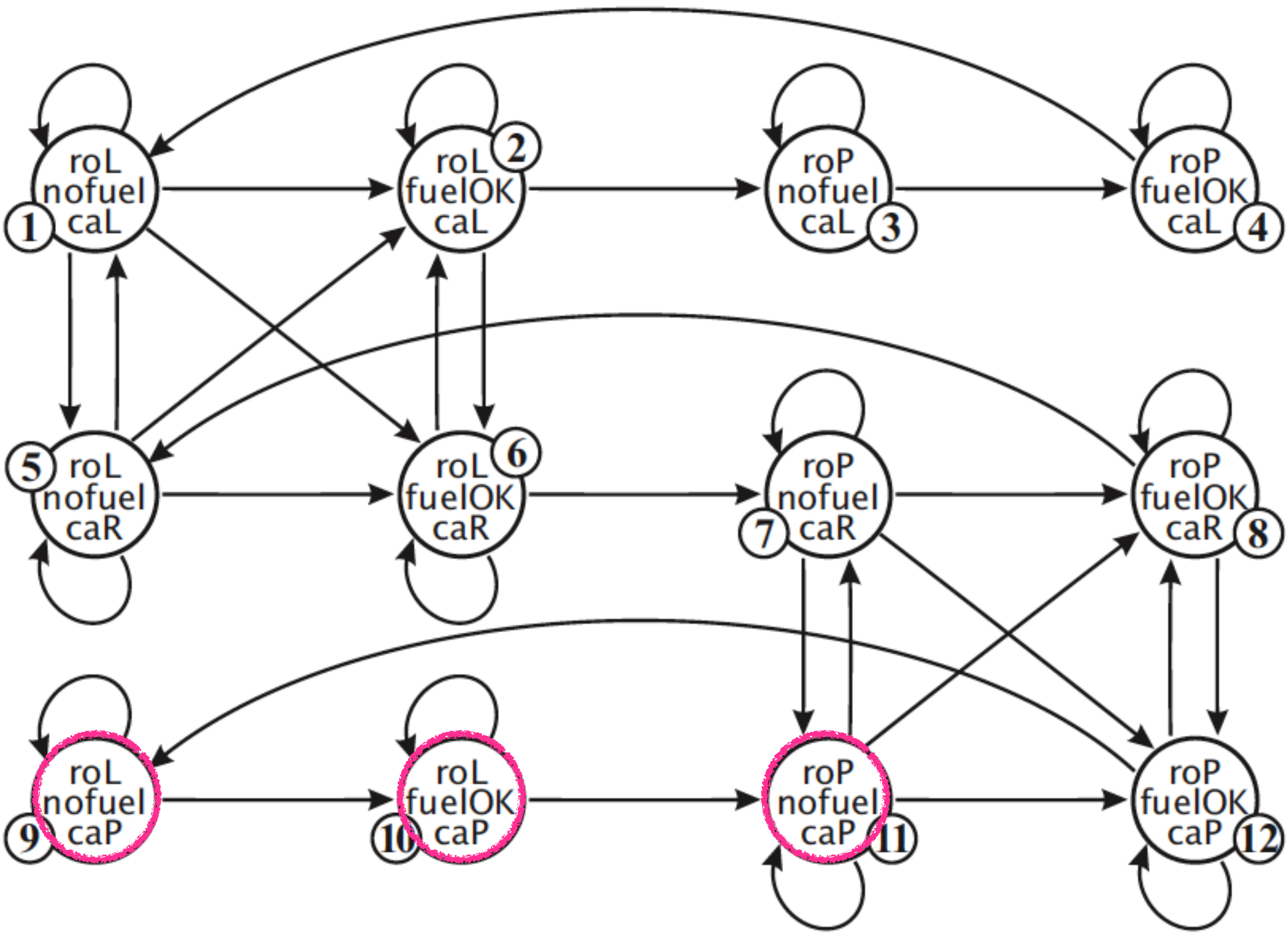
Example


roL
nofuel
caL
1
roL
fuelOK
caL
2
roP
nofuel
caL
3
roP
fuelOK
caL
4
5
roL
nofuel
caR
roL
fuelOK
caR
6
roP
nofuel
caR
7
roP
fuelOK
caR
8
roL
nofuel
caP
9
roL
fuelOK
caP
10
roP
nofuel
caP
11
roP
fuelOK
caP
12

Example


roL
nofuel
caL
1
roL
fuelOK
caL
2
roP
nofuel
caL
3
roP
fuelOK
caL
4
roL
nofuel
caR
5
roL
fuelOK
caR
6
roP
nofuel
caR
7
roP
fuelOK
caR
8
roL
nofuel
caP
9
roL
fuelOK
caP
10
roP
nofuel
caP
11
roP
fuelOK
caP
12

Example


roL
nofuel
caL
1
roL
fuelOK
caL
2
roP
nofuel
caL
3
roP
fuelOK
caL
4
5
roL
nofuel
caR
roL
fuelOK
caR
6
roP
nofuel
caR
7
roP
fuelOK
caR
8
roL
nofuel
caP
9
roL
fuelOK
caP
10
roP
nofuel
caP
11
roP
fuelOK
caP
12

Dynamic Logic

Dynamic Logic

1^{st} idea: Consider actions or programs α. Each such α defines a transition (accessibility relation) from worlds into worlds.

Dynamic Logic

1st idea: Consider actions or programs α. Each such α defines a transition (accessibility relation) from worlds into worlds.

2nd idea: We need statements about the outcome of actions:

- $[\alpha]\varphi$: "after every execution of α, φ holds,
- $\langle\alpha\rangle\varphi$: "after some executions of α, φ holds.

Dynamic Logic

1^{st} idea: Consider actions or programs α. Each such α defines a transition (accessibility relation) from worlds into worlds.

2^{nd} idea: We need statements about the outcome of actions:

- $[\alpha]\varphi$: "after every execution of α, φ holds,
- $\langle\alpha\rangle\varphi$: "after some executions of α, φ holds.

As usual, $\langle\alpha\rangle\varphi \equiv \neg[\alpha]\neg\varphi$.

Dynamic Logic

OI

3^{rd} idea: Programs/actions can be combined (sequentially, nondeterministically, iteratively), e.g.:

$$[\alpha;\beta]\varphi$$

would mean "after every execution of α and then β, formula φ holds".

Dynamic Logic

Definition 3.1 (Labelled Transition System)

A labelled transition system is a pair

$$\langle St, \{\xrightarrow{\alpha} : \alpha \in \mathbf{L}\}\rangle$$

where St is a non-empty set of states and $\mathbf{L}$ is a non-empty set of labels and for each $\alpha \in \mathbf{L}$: $\xrightarrow{\alpha} \subseteq St \times St$.

Dynamic Logic

Definition 3.1 (Labelled Transition System)

A labelled transition system is a pair

$$\langle St, \{\xrightarrow{\alpha} :\ \alpha \in \mathbf{L}\}\rangle$$

where St is a non-empty set of states and $\mathbf{L}$ is a non-empty set of labels and for each $\alpha \in \mathbf{L}$: $\xrightarrow{\alpha} \subseteq St \times St$.

Definition 3.2 (Dynamic Logic: Models)

A model of propositional dynamic logic is given by a labelled transition systems and an evaluation of propositions.

Dynamic Logic

Definition 3.2 (Dynamic Logic: Models)

A model of propositional dynamic logic is given by a labelled transition systems and an evaluation of propositions.

Dynamic Logic

01

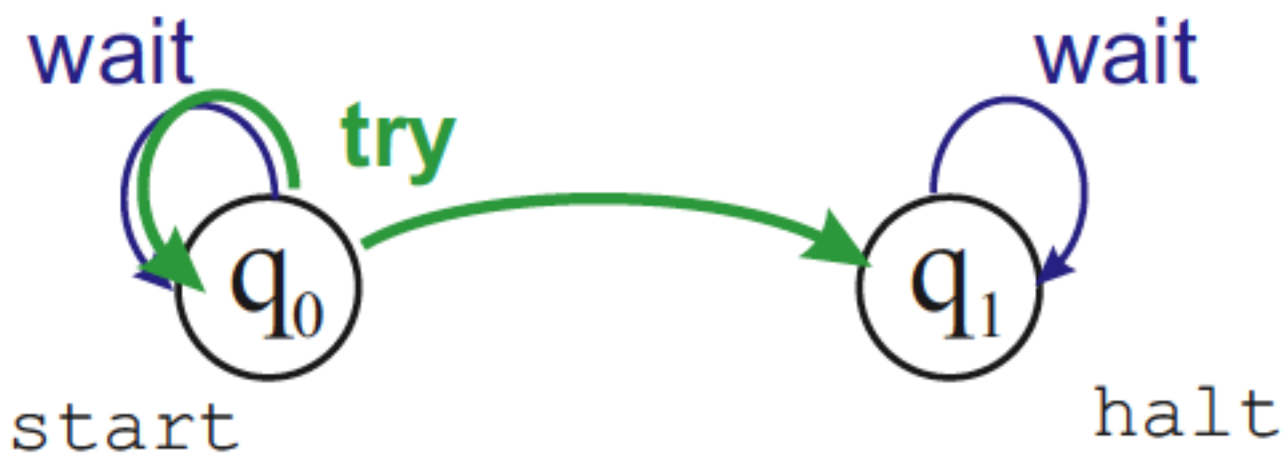
Definition 3.2 (Dynamic Logic: Models)

A model of propositional dynamic logic is given by a labelled transition systems and an evaluation of propositions.

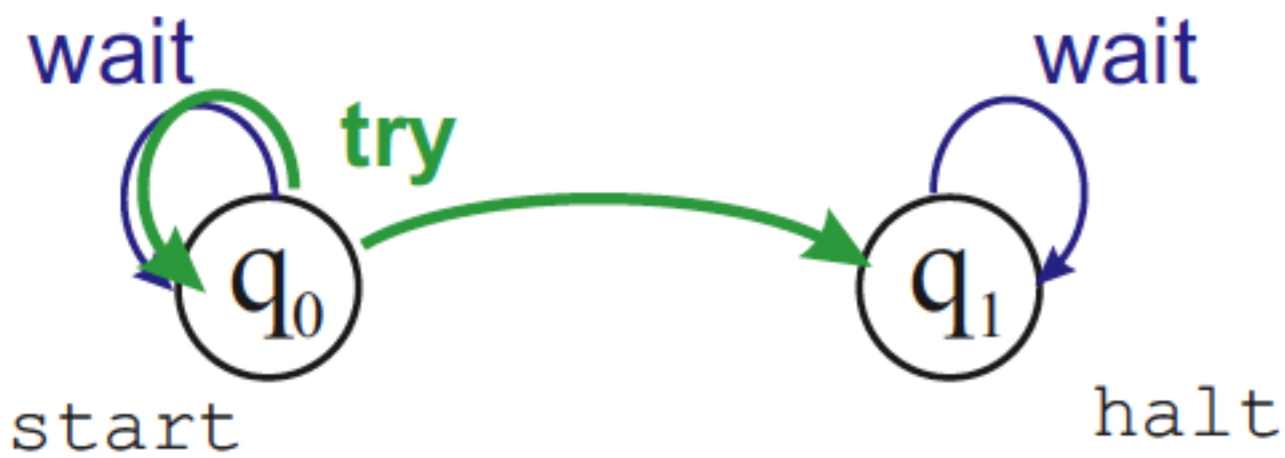
Definition 3.3 (Semantics of DL)

$\mathcal{M}, s \models [\alpha]\varphi$ iff for every t such that $s \xrightarrow{\alpha} t$, we have $\mathcal{M}, t \models \varphi$.

Dynamic Logic

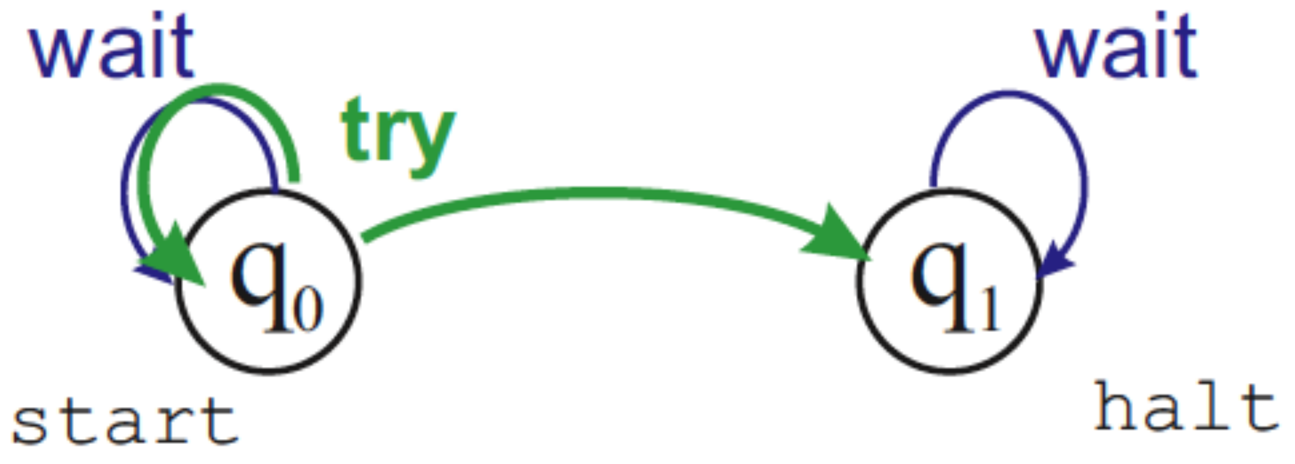


Dynamic Logic



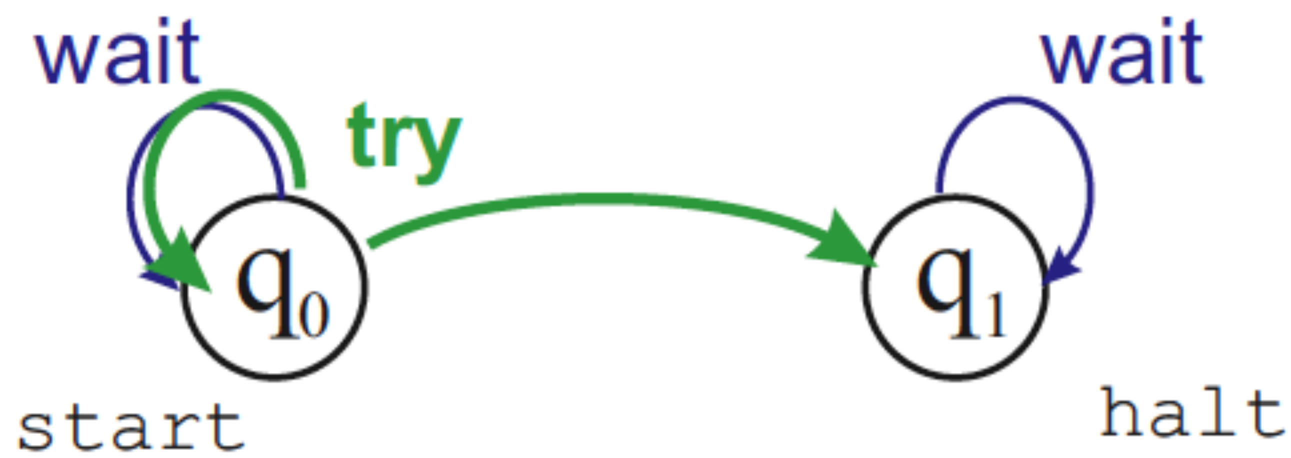
$$\text{start} \rightarrow \langle \mathit{try} \rangle \text{halt}$$

Dynamic Logic



$$\text{start} \longrightarrow \langle \mathit{try} \rangle \text{halt}$$
$$\text{start} \longrightarrow \neg [\mathit{try}] \text{halt}$$

Dynamic Logic



$$start \rightarrow \langle try \rangle halt$$

$$start \rightarrow \neg [try] halt$$

$$start \rightarrow \langle try \rangle [wait] halt$$

Concluding Remarks

- Practical Importance of Temporal and Dynamic Logics:
  - Automatic verification in principle possible (model checking).
  - Can be used for automated planning.
  - Executable specifications can be used for programming.

- Note:

  When we combine time and actions with knowledge (beliefs, desires, intentions, obligations...), we finally obtain a fairly realistic model of MAS.

Models of Practical Reasoning: BDI

Process of figuring out what to do. Practical reasoning is a matter of weighing conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires/values/cares about and what the agent believes (Bratman)

- computational model of human decision process oriented towards an action, based on models of existing mental models of the agents

- human practical reasoning consists of two activities:
  - deliberation: deciding what state of affairs we want to achieve and
  - means-ends reasoning (planning): deciding how to achieve these states

- the outputs of deliberation process are intentions

BDI Architecture

01

- BELIEFS
  - collection of information that the agents has about its the status of the environment, peer agents, self
- DESIRES
  - set of long term goals the agent wants to achieve
- INTENTIONS
  - agents immediate commitment to executing an action, either high-level or low level (depends on agents planning horizon)

- BDI architecture connects: (i) reactive (ii) planning & (iii) logical represention. BDI architecture does not count on theorem proving

$$\text{if } \varphi \in \mathcal{L}_{agent} \text{ then } \varphi, (\text{Bel } A\ \varphi), (\text{Des } A\ \varphi), (\text{Int } A\ \varphi) \in \mathcal{L}_{bdi}$$

BDI Inference Algorithm

- Basic algorithm:

1. initial beliefs → Bel
2. while true do
3.   Read(get_next_percept) → in
4.   Belief-revision(Bel, in) → Bel
5.   Deliberate(Bel, Des) → Int
6.   Plan(Bel, Int) → π
7.   Execute(π)
8. end while

BDI Modal Properties

- BELIEFS
  - KD45 system, modal logic where the B relation is serial, transitive and euclidean: satisfies K axioms, positive introspection axiom (4 axiom), negative introspection axiom (5 axiom), beliefs consistency axiom (D axiom).

- DESIRES
  - KD system, modal logic requiring desired goals not to contradict (D axiom).

$$(\text{Des } A\ \varphi) \rightarrow \neg(\text{Des } A\ \neg\ \varphi)$$

- INTENTIONS
  - KD system, modal logic requiring intentions not to contradict (D axiom).

$$(\text{Int } A\ \varphi) \rightarrow \neg(\text{Int } A\ \neg\ \varphi)$$

Properties of Intentions

- Intention persistency:
  - agents track the success of their intentions, and are inclined to try again if their attempts fail

$$(\mathsf{Int}\ A\ \varphi) \curvearrowleft \varphi$$

- Intention satisfiability:
  - agents believe their intentions are possible; that is, they believe there is at least some way that the intentions could be brought about.

$$(\mathsf{Int}\ A\ \varphi) \Rightarrow \mathsf{EF}\varphi$$

Properties of Intentions

- Intention-belief inconsistency:
  - agents do not believe they will not bring about their intentions; it would be irrational of agents to adopt an intention if believed was not possible

$$(\mathsf{Int}\ A\ \varphi) \wedge (\mathsf{Bel}\ A \neg \mathsf{EF}\varphi)$$

- Intention-belief incompleteness:
  - agent do not believe that their intention is possible to be achieved, may be understood as rational behavior

$$(\mathsf{Int}\ A\ \varphi) \wedge (\neg \mathsf{Bel}\ A \mathsf{EF}\varphi)$$

  - agents admit that their intentions may not be implemented.

$$(\mathsf{Int}\ A\ \varphi) \wedge (\mathsf{Bel}\ A\ \mathsf{EF}\neg\varphi)$$

Properties of Intentions

OI

- Intention side-effects:
  - Agents need not intend all the expected side effects of their intentions. Intentions are not closed under implication.

$$(\mathsf{Bel}\ A\ \psi \Rightarrow \varphi) \wedge (\mathsf{Int}\ A\ \psi) \wedge \neg(\mathsf{Int}\ A\ \varphi)$$

   * *is thus classified as fully rational behaviour*

  - Example: I may believe that going to the dentist involves pain, and I may also intend to go to the dentist - but this does not imply that I intend to suffer pain!

Rationality of Invetibilities & Options

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Rationality of Invetibilities & Options

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Rationality of Invetibilities & Options

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$

$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$

$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$

$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$

$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Agents Individual/Social Commitments

- Commitments: knowledge structure, declarative programming concept based on intentions (intentions are special kinds of comms).
  - specify relationships among different intentional states of the agents
  - specify social relations among agents, based on their comms to joint actions

The commitment is an agent's state of 'the mind' where it commits to adopting the single specific intention or a longer term desire.

- We distinguish between:
  - specific, commonly used comms — general comms
  - individual comms — social comms

Individual Commitments

- A can get committed to its intention φ in several different ways:
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft (\mathsf{Bel}\ A\ \varphi))$$

Individual Commitments

- A can get committed to its intention φ in several different ways:
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft (\mathsf{Bel}\ A\ \varphi))$$

  - single-minded commitment: besides above it intends the intention until it believes that it is no longer possible to achieve the goal

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft ((\mathsf{Bel}\ A\ \varphi) \vee (\mathsf{Bel}\ A\ \neg\mathsf{EF}\varphi))$$

Individual Commitments

- A can get committed to its intention φ in several different ways:
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft (\mathsf{Bel}\ A\ \varphi))$$

  - single-minded commitment: besides above it intends the intention until it believes that it is no longer possible to achieve the goal

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft ((\mathsf{Bel}\ A\ \varphi) \vee (\mathsf{Bel}\ A\ \neg\mathsf{EF}\varphi))$$

  - open-minded commitment: besides above it intends the intention as long as it is sure that the intention is achievable

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft ((\mathsf{Bel}\ A\ \varphi) \vee \neg(\mathsf{Bel}\ A\ \mathsf{EF}\varphi))$$

General Commitments

- Commitment is defined as $(\text{Commit } A\ \varphi\ \psi\ \lambda)$, where
- Convention is defined as $\lambda = \{\langle \rho_k, \gamma_k \rangle\}_{k \in \{1,\ldots,l\}}$
  - provided $\curvearrowleft$ stands for *until*, A stands for *always in the future*, Int is *agent's intention* and Bel is *agent's belief* then for $\lambda = \langle \rho, \gamma \rangle$ the commitment has the form:

$$(\text{Commit } A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\text{Int } A\ \varphi) \wedge \textit{decommitment\_rule} \qquad) \curvearrowleft \gamma)$$

$$(\text{Commit } A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\text{Int } A\ \varphi) \wedge ((\text{Bel } A\ \rho) \Rightarrow \mathsf{A}(\text{Int } A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

General Commitments

- Commitment is defined as $(\text{Commit } A\ \varphi\ \psi\ \lambda)$, where
- Convention is defined as $\lambda = \{\langle \rho_k, \gamma_k \rangle\}_{k \in \{1,\ldots,l\}}$
  - provided $\curvearrowleft$ stands for *until*, A stands for *always in the future*, Int is *agent's intention* and Bel is *agent's belief* then for $\lambda = \langle \rho, \gamma \rangle$ the commitment has the form:

$$
\begin{aligned}
(\text{Commit } A\ \varphi\ \psi\ \lambda) \quad \equiv \\
& \psi \wedge \mathsf{A}((\text{Int } A\ \varphi) \wedge \\
& \quad ((\text{Bel } A\ \rho_1) \Rightarrow \mathsf{A}(\text{Int } A\ \gamma_1)) \curvearrowleft \gamma_1) \\
& \quad \ldots \\
& \quad ((\text{Bel } A\ \rho_l) \Rightarrow \mathsf{A}(\text{Int } A\ \gamma_l)) \curvearrowleft \gamma_l) \\
& \curvearrowleft \bigvee_i \gamma_i)
\end{aligned}
$$

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

$$(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge ((\mathsf{Bel}\ A\ \rho) \Rightarrow \mathsf{A}(\mathsf{Int}\ A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

Joint (Social) Commitment

OI

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

$$(\text{Commit } A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\text{Int } A\ \varphi) \wedge ((\text{Bel } A\ \rho) \Rightarrow \mathsf{A}(\text{Int } A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

$$\begin{aligned}(\text{J-Commit } \Theta\ \varphi\ \psi\ \lambda) \quad \equiv \quad & \\ & \forall A :\ (A \in \theta) \Rightarrow \\ & \psi \wedge \mathsf{A}((\text{Int } A\ \varphi) \wedge \\ & \quad ((\text{Bel } A\ \rho) \Rightarrow \mathsf{A}(\text{Int } A\ \gamma) \curvearrowleft \gamma) \\ & \curvearrowleft \gamma)\end{aligned}$$

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)
  – for a convention in the form of $\lambda = \{\langle \rho_k, \gamma_k \rangle\}_{k \in \{1,\ldots,l\}}$

$$(\text{J-Commit}\ \Theta\ \varphi\ \psi\ \lambda) \equiv \forall A:\ (A \in \theta) \Rightarrow \psi \wedge \mathsf{A}((\chi_1 \wedge \chi_2) \curvearrowleft \chi_3)$$

where

$$\begin{aligned}
\chi_1 &= (\text{Int}\ A\ \varphi) \\
\chi_2 &= ((\text{Bel}\ A\ \rho_1) \Rightarrow \mathsf{A}((\text{Int}\ A\ \gamma_1) \curvearrowleft \gamma_1)) \wedge ((\text{Bel}\ A\ \rho_2) \Rightarrow \\
&\quad \mathsf{A}((\text{Int}\ A\ \gamma_2) \curvearrowleft \gamma_1) \wedge \cdots \wedge ((\text{Bel}\ A\ \rho_n) \Rightarrow \mathsf{A}((\text{Int}\ A\ \gamma_n) \curvearrowleft \gamma_n))) \\
\chi_3 &= \gamma_1 \vee \gamma_1 \vee \cdots \vee \gamma_n
\end{aligned}$$

Blind Social Commitment

01

- each agent is trying to accomplish the commitment until achieved

$$\lambda_{blind} = \{\langle(\text{Bel } A\ \varphi), (\text{M-Bel } \Theta\ \varphi)\rangle\}$$
$$\psi_{blind} = \neg(\text{Bel } A\ \varphi)$$

$$\begin{aligned}(\text{J-Commit } \Theta\ \varphi\ \psi\ \lambda) \quad \equiv \quad & \forall A:\ (A \in \Theta) \Rightarrow \\ & (\neg(\text{Bel } A\ \varphi) \wedge (\text{A}((\text{Int } A\ \varphi) \wedge \\ & \quad ((\text{Bel } A\ \varphi) \Rightarrow \text{A}((\text{Int } A\ (\text{M-Bel } \Theta\ \varphi)) \\ & \quad \curvearrowleft (\text{M-Bel } \Theta\ \varphi)))) \\ & \curvearrowleft (\text{M-Bel } \Theta\ \varphi))).\end{aligned}$$

Minimal Social Commitment

01

- minimal social commitment, also related to as joint persistent goal:
  - initially agents do not believe that goal is true but it is possible
  - every agent has the goal until *termination condition* is true
  - until termination: if agent beliefs that the goal is either *true or impossible* than it will want the goal that it becomes a mutually believed, but keep committed
  - the termination condition is that it is *mutually believed* either goal is true or impossible to be true.

$$\psi_{soc} = \neg(\mathsf{Bel}\ A\ \varphi) \wedge (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$

$$\lambda_{soc} = \left\{ \begin{array}{l} \langle(\mathsf{Bel}\ A\ \varphi), (\mathsf{M\text{-}Bel}\ \Theta\ \varphi)\rangle, \\ \langle(\mathsf{Bel}\ A\ \mathsf{AG}\neg\varphi), (\mathsf{M\text{-}Bel}\ \Theta\ \mathsf{AG}\neg\varphi)\rangle \end{array} \right\}$$

Minimal Social Commitment

$$(\text{J-Commit}\ \Theta\ \varphi\ \psi_{soc}\ \lambda_{soc}) \equiv$$

$$\forall A,\ A \in \Theta : [\neg(\text{Bel}\ A\ \varphi) \wedge (\text{Bel}\ A\ \text{EF}\varphi)] \wedge$$

$$\text{A}\left[\left(\begin{array}{l}(\text{Int}\ A\ \varphi) \wedge \\ ((\text{Bel}\ A\ \varphi) \Rightarrow \text{A}((\text{Int}\ A(\text{M-Bel}\ \Theta\ \varphi))) \curvearrowleft \chi \wedge \\ ((\text{Bel}\ A\ \text{AG}\neg\varphi) \Rightarrow \text{A}((\text{Int}\ A(\text{M-Bel}\ \Theta\ \text{AG}\neg\varphi))) \curvearrowleft \chi\end{array}\right) \curvearrowleft \chi\right]$$

where $\chi \equiv ((\text{M-Bel}\ \Theta\ \varphi) \vee (\text{M-Bel}\ \Theta\ \text{AG}\neg\varphi)))$

Mutual Belief ?

Definition 1:

$$(\text{M-Bel}\ \Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{Bel}\ A\ (\text{M-Bel}\ \Theta\ \varphi))$$

Definition 2:

$$(\text{E-Bel}^{0}\ \Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{Bel}\ A\ \varphi)$$

$$(\text{E-Bel}^{k}\ \Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{E-Bel}^{k-1}\ \Theta\ \varphi)$$

$$(\text{M-Bel}\ \Theta\ \varphi) \equiv \forall m \in N\ (\text{M-Bel}^{m}\ \Theta\ \varphi)$$