Knowledge and some real-life problems

Knowledge in Multi-Agent Systems (MAS)

- **MAS** is any collection of interacting agents ("muddy children", poker game, distributed system running according to some protocol, …)
- Description of MAS has to focus on few details the agents need to work with when following the protocol – this is the **language** used by the agents
- At any point of time each agent is in some state – agent's **local state**

Local state of the poker player (Ag1) :
- Cards Ag1 is holding.
- Bits made by the other players.
- Any cards Ag1 has seen during the game.
- Additional information (e.g. the strategy of some other players, namely A likes to bluff, P is conservative, …)

Knowledge in Multi-Agent Systems (MAS)

- At any point of time the whole MAS is in some state – system's **global state**. How this can be represented? Can it be a list of local states of all agents?

> In case of ***a message passing system*** it is important to know how the communication channels work. This information is not provided in local state of any agent!
> We need to consider **environment**, too - it describes everything relevant what is not part of local states of agents..

- The system's **global state** is a list of local states of all the agents + local state of the environment, namely

$$\langle s_e, s_1, s_2, \ldots, s_n \rangle$$

- Time development and changes of the system are described as a function from time to $\mathcal{G}$ (set of all global states) and referred to as a **run**.

Knowledge in Multi-Agent Systems (MAS)

- ***Terminology***: Let **r** be a run and **m** a specific time, then the **point (r,m)** is the global state **r(m)=**<s_e, s_1,s_2,…, s_n> and **r**$_i$ **(m)=** s_i is the local state of the agent ***I,***

  r$_i$ **(m)=** s_e is the local state of the environment.

- A MAS system $\mathcal{R}$ over the set $\mathcal{G}$ of global states is a set of runs over $\mathcal{G}$.

  In practice, the appropriate set of runs is chosen or specified by the system designer.

- For using the language **£** in the **system** $\mathcal{R}$ we need to have its interpretation for any global state. This is provided by an interpreted system $\mathcal{I}$ consisting of a pair ($\mathcal{R}$,π), where the function π: $\mathcal{G}$ X **£** → **{true,false}**

Knowledge in Multi-Agent Systems (MAS)

- ***Terminology***: Let us consider two global states

  $\mathbf{s} = <s_e, s_1, s_2, \ldots, s_n>$ and

  $\mathbf{w} = <w_e, w_1, w_2, \ldots, w_n>$.
- The global states **s** and **w** are **not distinguishable** for the agent **i** (denoted as $\approx_i$) if this agent's local state in both **s** and **w** is identical, ie. $s_i = w_i$,

- Knowledge in an interpreted system $\mathcal{I} = (\mathcal{R}, \pi)$ with ***n*** agents excluding the environment, is defined through a Kripke structure ($\mathbf{S}$, π, $K_1, K_2, \ldots, K_n$), where
  - **S** is the set of all points in the interpreted systém $\mathcal{I}$ and
  - the admissibility relation K_i is specified by $\approx_i$

Knowledge base

is a system *gradually receiving information about its environment* that is supposed to *answer some queries*.

Let ***the knowledge base KB*** *be one of the agents in a larger system*, where at least the following agents are present:

- ***Environment*** is an agent representing the model of the external world.
- ***Teller*** is the agent telling the ***KB*** the facts about the external world.

Environment´s state is expected to contain complete description (of the relevant features) of the external world.

Local state of ***KB*** describes information the ***KB*** has obtained about the external world up to now.

Local state of ***Teller*** contains all he knows about „the external world“ + „the knowledge up to now provided to ***KB***“ ...

This informal description of ***KB*** is still rather vague – it offeres many possibilities for modelling the global states.

The simplest choice could be based on the following constraints:

• The external world can be described using a set of primitive propositions Φ.

• The external world is ***stable***, ie. the truth values of the primitive propositions about the external world do not change in time.

• The *Teller* has ***full information*** about the external world.

•All information *stored in* ***KB*** is ***valid***.

• There is *no a priori knowledge* about the external world or about the content of ***KB***.

For simplicity let us assume, that

• The **external world** can be described by an evaluation **α** of primitive propositions from **Φ** (and this evaluation **α** remains fixed!).

• The **Teller** has access to the evaluation **α** and to the sequence of facts, he has provided as input into ***KB*** so far.

• ***KB´s* local state** contains the sequence $A_1, \ldots, A_n$ of formulas provided by the **Teller** as input (the formulas A_i could be either ***propositional*** *or* ***modal formulas***).

• **Global state** is denoted by $(\alpha, \langle A_1, \ldots, A_n \rangle, .)$

1.

KB that stores **propositional formulas, only!**
ie. ***KB*** stores ***propositional formulas*** created from the set **Φ** of primitive propositions**.**

- The ***queries*** posed to (subjected to) ***KB*** have the form of propositional formulas created from propositions **Φ .**
- All formulas *stored* in ***KB*** *are true* under the evaluation **α.**
- There are no *a priori assumptions* about the external world and about information provided by the ***Teller*** to ***KB***.

These assumptions represent the constraints that will be used to construct the corresp. Kripke str.

Let I^{kb} describe gradual addition of information into the ***KB*** in the form of a sequence r of the global states. Let us define

$(I^{kb}, r, m) \models K_{KB}\varphi$ iff φ holds in **all the time points** ***KB*** cannot distinguish from (r, m) (they represent„summary of what ***KB*** has been told up to now").

How the *KB* should answer a query *B* ?

Assuming that ***KB*** remembers **conjunction of all information provided up to now**.

Choice **a)**

Suppose ***KB*** is asked at a certain moment (r,m) the query ***B***, where ***B*** is a propositional formula. Our knowledge base has no access to the external world – that is why ***B*** cannot be interpreted as a query about the external world: ***to be fully correct*** one should interpret the query ***B*** as a question „What the ***KB*** knows about the external world, ie. $K_{KB}(B)$**?** ".

The KB should answer this question as follows

$$\begin{cases} \text{YES} & \text{if } (I^{kb}, r, m) \models K_{KB}B \\ \text{NO} & \text{if } (I^{kb}, r, m) \models K_{KB}\neg B \\ \text{Do Not Know} & \text{else} \end{cases}$$

Choice b)

Suppose ***KB*** is in the local state

$\langle A_1, \ldots, A_k \rangle$, denote $\kappa = A_1 \wedge \ldots \wedge A_k$

and let the knowledge base know everything what is a consequence of κ.

In such a case **the query *B*** can get a positive answer in 2 cases:

$$\text{YES iff } \begin{cases} B & \text{is a consequence of } \kappa \\ \text{or} & \\ K_{KB}B & \text{is a consequence of } K_{KB}\kappa \end{cases}$$

Analogous solution for the negative answer to the query ***B***:

$$\text{NO iff } \begin{cases} \neg B & \text{is a consequence of } \kappa \\ \text{or} & \\ K_{KB}\neg B & \text{is a consequence of } K_{KB}\kappa \end{cases}$$

DoNotKnow in all remaining case.

*If **KB** contains only propositional formulas about the external world* and no facts about ***KB*** itself, then the answers to propositional queries are the same for both choices a) and b):

KB theorem 1.

Let us assume that KB contains propositional formulas only, ie.

$$r_{KB}(m) = \langle A_1, \ldots, A_k \rangle, \quad \kappa = A_1 \wedge \ldots \wedge A_k$$

and B is a propositional formula.

In this case the following are **equivalent**

$$(i)\ (\mathcal{I}^{kb}, r, m) \models K_{KB} B$$

$$(ii)\ \kappa \rightarrow B \quad \text{is a propositional tautology}$$

$$(iii)\ M_n^{rst} \models K_{KB}\kappa \rightarrow K_{KB} B$$

laboratory Gerstner

2.

What happens if the query is not limited to **propositional formulas ?**

Let us consider a query $B \equiv (p \rightarrow K_{KB}\, p)$

"Is it the case, that if p holds then KB knows it?"

How the answer for the query ***B*** should be obtained? Here, it seems reasonable to apply the choice a) used earlier which suggests the answer to ***B*** as follows:

$$\begin{cases} \text{YES} & \text{if } (\mathcal{I}^{kb}, r, m) \models K_{KB} B \\ \text{NO} & \text{if } (\mathcal{I}^{kb}, r, m) \models K_{KB} \neg B \\ \text{Do Not Know} & \text{else} \end{cases}$$

What do we know about validity of $K_{KB}(p \rightarrow K_{KB}\, p)$? (1)

According to **Mod_T8 a, Mod_T8b** we know that $K_{KB}\,(p \rightarrow K_{KB}\, p\,)$ is equivalent to the formula $K_{KB}\, p \lor K_{KB} \neg\, p$ (2)

Now, we are ready to answer the query $\boldsymbol{B = (p \rightarrow K_{KB}\, p)}$ by considering the equivalent transcriptions

1. $\boldsymbol{K_{KB}\, (p \rightarrow K_{KB}\, p)} = K_{KB}\, p \vee K_{KB} \neg p$ (see Mod_T8)
2. $\boldsymbol{K_{KB} \neg (p \rightarrow K_{KB}\, p)} = K_{KB}\, (p \;\&\; \neg K_{KB}\, p)$

We have proven earlier in **Mod_T7** that the formula $K_{KB}\, (p \;\&\; \neg K_{KB}\, p)$ is contradictory (and thus it cannot be proven)!
Consequently, the query ***B*** can never result in the answer NO!

To answer the query $\boldsymbol{B = (p \rightarrow K_{KB}\, p)}$ let us consider the equivalent transcription $\boldsymbol{K_{KB}\, (p \rightarrow K_{KB}\, p)} = K_{KB}\, p \vee K_{KB} \neg p$. Thus the answer is

YES	if p or $\neg p$ is a consequence of KB content
Do Not Know	else.

3.

Does it make sense to provide *KB* with input described by non-propositional formulas ?

Let us assume, that ***KB*** has current information represented by a sequence of formulas $<F_1, \ldots, F_i>$ and the next provided information is $\boldsymbol{F_{i+1} = (p \rightarrow K_{KB}\, p)}$ saying „if ***p*** is valid, ***KB*** knows it".

This information can be useful, if ***KB*** can verify what it knows and what it does not know. If there holds $<F_1, \ldots, F_i> \vdash \neg K_{KB}\, p$, then ***KB*** can deduce from the input $\boldsymbol{F_{i+1}}$ that p does not hold.

This shows that thanks to the non-propositional input the ***KB*** can gain (derive) new information about the external world using **introspection.**

If KB obtains non-propositional input, **its knowledge can** no more be represented as a ***conjunction of the input sequence*** (as in the propositional case).

Let us explain this using an ***example situation***, when the provided input represents a claim, that

- has been true at the moment of input,
- does not remain true in any later time point.

Under such conditions it is impossible to describe knowledge in the form of ***conjunction of provided input***. ***Why?***

We will show an example in which the considered conjunction results in a contradiction (from which anything can be derived) !

Example: Assume ***KB*** obtains as input a fact, that

- is true in the current moment,
- but it does not remain valid in any possible future state.

Let the primitive proposition ***p*** about the external world be valid, but the ***KB*** does not know ***p*** yet (until the tact ***j*** no such information has been provided). In such a situation there must hold

$$\boldsymbol{p} \,\&\, \neg K_{KB}\boldsymbol{p} \qquad (3)$$

This can be provided by the *Teller* in the tact ***j* +1**. But even given this information (3), we cannot claim that ***KB*** knows (3), ie.the formula $K_{KB}(p \,\&\, \neg K_{KB}\, p)$ holds in the next time point. **Why? This is impossible** since this formula contradicts **S5** ie.($\mathbf{K_n}$+Ax3+Ax4+Ax5).

But even in this case, the ***KB*** should gain some knowledge, namely $K_{KB}\boldsymbol{p}$ should hold after the ***KB*** has been told $\boldsymbol{p} \,\&\, \neg K_{KB}\boldsymbol{p}$

$$\boldsymbol{p} \,\&\, \neg K_{KB}\, \boldsymbol{p} \qquad (3)$$

We have learned already through the analysis of the formula (3) that ***obtaining*** information φ **does not have to result** in a situation when the ***KB*** knows it (there holds $K_{KB}\,\varphi$).

Intuitively, the ***KB*** should gain some knowledge after it has been told (3) : it should know, that ***p*** is true! Consequently input of (3) should result in $K_{KB}\,\boldsymbol{p}$.

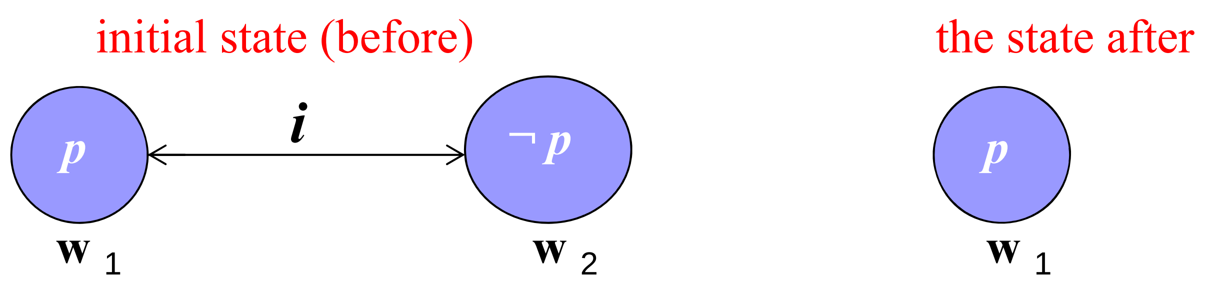
How to make this intuition true? ***The knowledge base cannot treat modal formulas in the same way as the propositional formulas***!

A new approach has to be designed. The viable solution is based on **modal logics**!

Common knowledge

Does a **public announcement** result in common knowledge ever ?

- Suppose you tell to Mr.***i*** : „*You do not know it, but I have to tell you that you have some bug on your back.*“
- Let us abbreviate the claim „*You have some bug on your back* .“ by ***p*** and let γ for the formula $\neg K_i p \,\&\, p$



In the initial state it is obvious, that $K_i \gamma$ does not hold in the state w_1. Consequently, it cannot be the case that $C\gamma$ holds initially! On the other hand $C\gamma$ holds in the state „after“.

Common knowledge in ASMP (asynchronous message passing systems)

Asychronous sending of messeges

- Let us denote $\boldsymbol{\alpha}$ a fact that describes agent´s 1 knowledge in time 0: „*Agent* 2 *did not obtain any message from me, yet*.“ It must be the case that $K_1 \boldsymbol{\alpha}$ holds in the time 0.
- Agent 1 sends a message „*Hello*.“ to the agent 2.
- This act of sending a message results in a **loss** of certain knowledge, namely $K_1 \alpha$ cannnot be true any more !

It can be shown that there is a close connection between common knowledge and synchronization of agent's activities!

Communication and common knowledge

Let us consider 2 generals, who communicate by sending letters through human messengers: When can they be sure that both their armies can attack jointly (in coordinated way)?

Common knowledge is closely related to communication means used in the system. It can be proven, that in an asynchronous system no new knowledge or common knowledge can be gained through **sending messeges**:

Theorem: For any structure ***S*** and any formula ψ there holds that during in any time development of this system (described by a run ***r***) and at any instant ***m*** :

$$(S,r,m) \models C_G\psi \text{ if and only if } (S,r,0) \models C_G\psi$$

Recommended resources

- **Modal logic**

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, Moshe Y. Vardi: *Reasoning About Knowledge*, MIT Pres 1995, 2003

- **Temporal logic**

Michael Fisher: *Introduction into Formal Methods Using Temporal Logic*, John Wiley & Sons, 2011

- **Modal and temporal logics**

Chapter 8 in the volume *Umělá inteligence*(6), Academia 2012

FIRST PhD Autumn School on Modal Logic, November 10-11 2009, materials from the courses *Temporal Logics for Specification and Verification* (V. Goranko), *Computational Modal Logic* (C. Areces, P. Blackburn), http://hylocore.ruc.dk/m4m6school.html