ESSAI 2024 course: Logic-based specification and verification of multi-agent systems

Lecture 2: Logics for temporal strategic reasoning with incomplete and imperfect information

Valentin Goranko
Stockholm University

2nd European Summer School on Artificial Intelligence
ESSAI 2024
Athens, July 15-19, 2024

Outline

- Modelling strategic reasoning with incomplete and imperfect information.
- ATL with incomplete/imperfect information.
- Epistemic extensions of ATL.
- An application to multi-agent planning.
- Concluding remarks.

Strategic reasoning with incomplete information

Incomplete information vs imperfect information

The decision making and abilities of strategically reasoning players crucially depend on the knowledge they possess about the game/system, other players' abilities and goals, etc.

So far I have considered structures of *complete and (almost) perfect information*. In reality this is seldom the case.

Thus, the question arises:
what can players achieve in a game/MAS if they are not completely informed about its structure and the current play?

Note the distinction:

– incomplete information: about the game structure, rules, players' possible actions, etc.
– imperfect information: about the play (current state, players' actions, etc

Incomplete information usually implies imperfect information.

On the other hand, any game of incomplete information can be regarded as a game of imperfect information that starts with Nature choosing a game in a way that players are not perfectly informed about. (Harsanyi's reduction)

Hereafter, both terms will be used almost interchangeably.

Knowledge and strategic abilities of agents

Consider:

- The agent **a** has a strategy to eventually achieve the goal γ
- The agent **a** knows that she has a strategy to eventually achieve γ.
- The agent **a** knows a strategy to achieve γ

Clearly, these are different, and only the last version implies practical ability.

Knowledge and strategies of coalitions

For coalitions, things become even more complicated. Compare:

- The coalition A has a joint strategy to achieve γ.
- Every agent in the coalition A knows that the coalition has a joint strategy to achieve γ.
- It is a common knowledge in the coalition A that it has a joint strategy to achieve γ.
- Every agent in A knows a joint strategy to achieve γ.
- A joint strategy to achieve γ is a common knowledge in A.

Can any of these guarantee that a coalition of rational players can achieve the goal γ?

The coordinated attack problem revisited

Two armies, positioned at the opposite sides of a castle intend to attack the common enemy in the castle.

They can only succeed if they attack together and simultaneously.

The armies have two choices: to attack at dawn or to attack at dusk tomorrow.

In order to coordinate the attack, the army generals must exchange messages via messenger. However, he can be captured by the enemy on his way there, or on his way back, or ...

Thus, it can be proved that coordination (i.e., common knowledge of the time of the planned attack) in this situation is impossible.

Describing the coordinated attack problem in ATL

The armies: A_1, A_2; ‘coordination’: C; ‘victory’: V.

It is known that $\neg\langle\langle A_1, A_2\rangle\rangle \mathcal{F} C$ holds.

But: $\neg\langle\langle A_1, A_2\rangle\rangle \mathcal{F} C \to \neg\langle\langle A_1, A_2\rangle\rangle \mathcal{F} V$ should be assumed valid.

These together lead to the conclusion that $\neg\langle\langle A_1, A_2\rangle\rangle \mathcal{F} V$.

However, intuitively the coalition $\{A_1, A_2\}$ does have a strategy to win, e.g., by both armies attacking simultaneously at 8am.

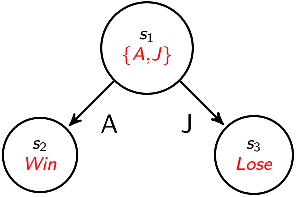
So, is $\langle\langle A_1, A_2\rangle\rangle \mathcal{F} V$ true, after all?

It depends on whether they can coordinate.

The ace and joker game

Consider the following game:

Two cards, **Ace** and **Joker**, lie face down. The player must choose one. The Ace wins, the Joker loses.



Does the player have a strategy to win the game?

Does the player know that she has a strategy to win the game?

Does the player know a strategy to win the game?

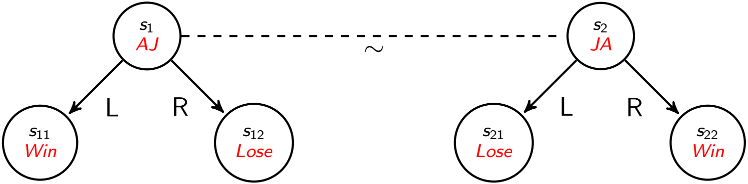
Again, does the player have a strategy to win the game?

It depends on what 'strategy' in the case of incomplete information means.

The game model above is incorrect and misleading!

A concurrent game model with incomplete information modelling the Ace and Joker game

There are two possible initial states, not one! They are *AJ* and *JA*.
They lead to two different game trees:



The player cannot distinguish states s_1 and s_2.

Concurrent game models with incomplete information (CGMII):
add an indistinguishability relation on states for each agent.

NB: indistinguishable states for an agent must enable the same actions for that agent.

Concurrent game models with incomplete information

Concurrent game models with incomplete information (CGMII):

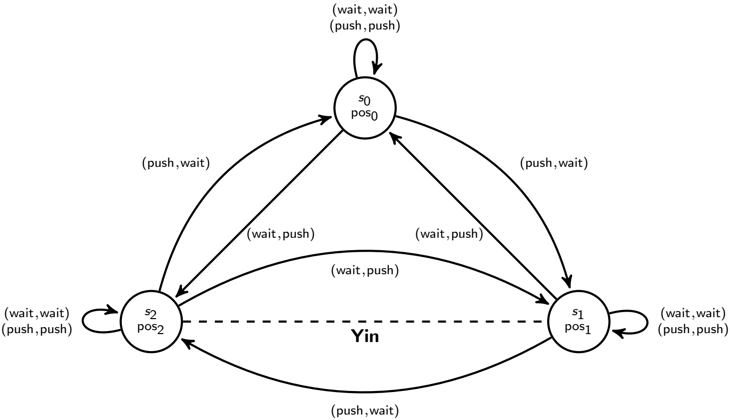
$$\langle \mathbb{A}, S, \{\sim_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{A}}, \mathsf{Act}, \mathsf{act}, \mathsf{out}, \mathsf{AP}, \mathsf{L} \rangle$$

where:

- $\langle \mathbb{A}, S, \{\sim_{\mathbf{i}}\}_{\mathbf{i}\in\mathbb{A}}, \mathsf{AP}, \mathsf{L} \rangle$ is a multi-agent epistemic model.
- $\langle \mathbb{A}, S, \mathsf{Act}, \mathsf{act}, \mathsf{out}, \mathsf{AP}, \mathsf{L} \rangle$ is a concurrent game model.
- For every $s_1, s_2 \in S$ and $\mathbf{i} \in \mathbb{A}$ such that $s_1 \sim_{\mathbf{i}} s_2$ it holds that

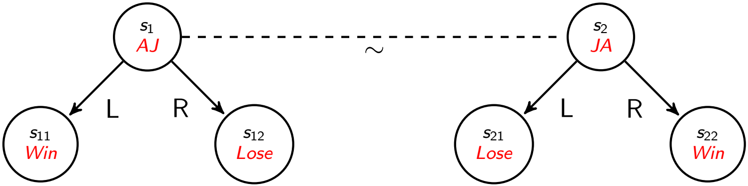
$$\mathsf{act}(\mathbf{i}, s_1) = \mathsf{act}(\mathbf{i}, s_2).$$

Concurrent game models with incomplete information: a variation of the two-robots example



Yin cannot distinguish states s_1 and s_2.

Uniform strategies



The problem with the Ace-Joker game is that the strategy "choose the Ace" is not executable in this game!

For a strategy to be executable by a player with imperfect information, it must be uniform: one that prescribes the same actions at indistinguishable states.

The player does not have a uniform strategy to ensure winning at both s_1 and s_2.

Strategic abilities in CGM with incomplete info: uniform memory-based strategies

The notion of uniformity extends to memory-based strategies.

A uniform memory-based strategy for a given player in a CGMII is one that prescribes the same actions for the player at any two histories that are indistinguishable for that player.

This extends in various ways to *uniform memory-based joint strategies* for coalitions. The simplest is to consider tuples of individually uniform strategies.

The semantics of $\langle\langle C \rangle\rangle$ in CGMII is adjusted accordingly:

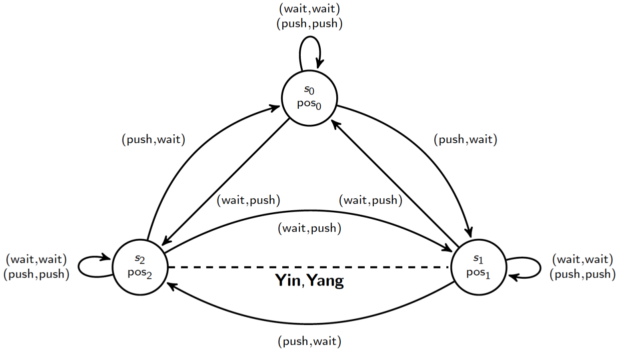
$\mathcal{M}, q \models_u \langle\langle C \rangle\rangle\gamma$ iff there is a uniform for every player in C joint strategy for C such that $\mathcal{M}, \lambda \models \gamma$, for every play λ consistent with that strategy.

More on that coming soon.

Subjective and objective abilities of players with imperfect information

- objective ($\models^o_u$): required to work only from the actual state;
- subjective ($\models^s_u$): required to work only from every state that is indistinguishable for the player from the actual state.
  - The blind driver scenario
  - Two-player team's strategy for exiting a maze
  - Student-supervisor scenario
- How to define objective and subjective abilities for coalitions? It depends on their abilities to communicate.

Strategic abilities with incomplete information some examples



$\mathcal{M}, s_1 \overset{?}{\models}{}^{o}_{u} \langle\!\langle \mathbf{Yin}, \mathbf{Yang} \rangle\!\rangle \mathcal{X}\, \mathsf{pos}_1$ Yes $\quad \mathcal{M}, s_1 \overset{?}{\models}{}^{s}_{u} \langle\!\langle \mathbf{Yin}, \mathbf{Yang} \rangle\!\rangle \mathcal{X}\, \mathsf{pos}_1$ No

$\mathcal{M}, s_1 \overset{?}{\models}{}^{s}_{u} \langle\!\langle \mathbf{Yin}, \mathbf{Yang} \rangle\!\rangle \mathcal{X} \langle\!\langle \mathbf{Yin}, \mathbf{Yang} \rangle\!\rangle ((\neg \mathsf{pos}_1)\, \mathcal{U}\, \mathsf{pos}_2)$ Yes, with memory

$\mathcal{M}, s_1 \overset{?}{\models}{}^{o}_{u} \langle\!\langle \mathbf{Yang} \rangle\!\rangle \mathcal{G}\, \neg \mathsf{pos}_0$ No (unless **Yang** can observe **Yin**'s actions)

Semantic variants of ATL with incomplete info and memory

Following Schobbens'2004:

- $\models_{IR}$: complete information and memory-based strategies;
- $\models_{Ir}$: complete information and memoryless strategies;
- $\models_{iR}$: incomplete information and memory-based strategies;
- $\models_{ir}$: incomplete information and memoryless strategies.

The semantic clause for $\models_{iR}$ is adjusted respectively, e.g.:

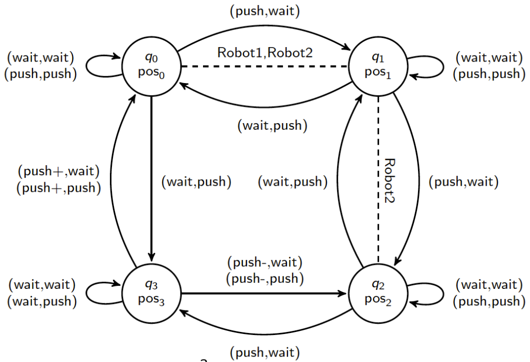
$\mathcal{M}, h \models_{iR} \langle\langle C \rangle\rangle\gamma$ iff there is a uniform memory-based joint strategy σ_C for C such that $\mathcal{M}, \lambda \models \gamma$, for every $\lambda \in \bigcup_{h' \in \mathsf{St} \text{ s.t.} h \sim_C h'} \mathsf{out}(h', \sigma_C)$, where

$\sim_C := \bigcup_{\mathbf{a} \in C} \sim_{\mathbf{a}}$ represents the *group knowledge of coalition C*

and $h \sim_C h'$ iff $|h| = |h'|$ and $h[i] \sim_C h'[i]$ for each $i \leq |h|$.

This captures *subjective abilities* of players with incomplete info.

Strategic abilities with incomplete info and memory: exercises



$\mathcal{M}, q_1 \overset{?}{\models}_{ir} \langle\langle 1,2 \rangle\rangle \mathcal{X}\, pos_2$ No $\quad \mathcal{M}, q_1 \overset{?}{\models}_{ir} \langle\langle 1,2 \rangle\rangle \mathcal{F}\, pos_2$ Yes

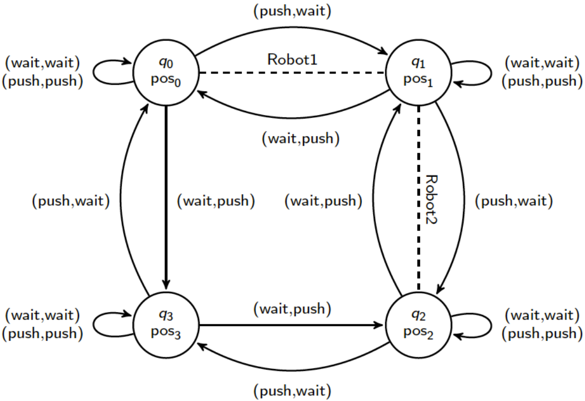
$\mathcal{M}, q_1 \overset{?}{\models}_{ir} \langle\langle 1 \rangle\rangle \mathcal{G}\, \neg pos_3$ Yes $\quad \mathcal{M}, q_0 \overset{?}{\models}_{ir} \langle\langle 2 \rangle\rangle \mathcal{G}\, \neg pos_2$ No

$\mathcal{M}, q_1 \overset{?}{\models}_{iR} \langle\langle 1 \rangle\rangle \mathcal{G}\, \neg pos_0$ No $\quad \mathcal{M}, q_0 \overset{?}{\models}_{ir} \langle\langle 1 \rangle\rangle \mathcal{G}\, \neg pos_2$ Yes

$\mathcal{M}, q_3 \overset{?}{\models}_{ir} \langle\langle 2 \rangle\rangle \mathcal{G}\, \neg pos_1$ No $\quad \mathcal{M}, q_3 \overset{?}{\models}_{iR} \langle\langle 2 \rangle\rangle \mathcal{G}\, \neg pos_1$ No

NB: one answer is wrong. Which one?

Strategic abilities with incomplete info and memory: more exercises



$\mathcal{M}, q_1 \overset{?}{\models}_{ir} \langle\langle 1,2 \rangle\rangle \mathcal{X}\, \mathsf{pos}_2$ Yes $\mathcal{M}, q_1 \overset{?}{\models}_{iR} \langle\langle 2 \rangle\rangle \mathcal{F}\, \mathsf{pos}_2$ No

$\mathcal{M}, q_1 \overset{?}{\models}_{iR} \langle\langle 1 \rangle\rangle \mathcal{G}\, \neg \mathsf{pos}_0$ No $\mathcal{M}, q_0 \overset{?}{\models}_{ir} \langle\langle 1 \rangle\rangle \mathcal{G}\, \neg \mathsf{pos}_2$ Yes

$\mathcal{M}, q_1 \overset{?}{\models}_{iR} \langle\langle 1,2 \rangle\rangle ((\neg \mathsf{pos}_0)\ \mathcal{U}\ \mathsf{pos}_2)$?

(Addendum) Comparing the semantics of strategic ability

W. Jamroga and N. Bulling, Comparing variants of strategic ability: how uncertainty and memory influence general properties of games, JAAMAS, 2014.

1. $\mathrm{ATL}[\mathit{ir}] \subsetneq \mathrm{ATL}[\mathit{Ir}]$.

   $\models_{\mathit{Ir}} (\varphi \vee \langle\langle C \rangle\rangle \mathcal{X} \langle\langle C \rangle\rangle \mathcal{F} \varphi) \leftrightarrow \langle\langle C \rangle\rangle \mathcal{F} \varphi$ but
   $\not\models_{\mathit{ir}} (\varphi \vee \langle\langle C \rangle\rangle \mathcal{X} \langle\langle C \rangle\rangle \mathcal{F} \varphi) \leftrightarrow \langle\langle C \rangle\rangle \mathcal{F} \varphi$

2. $\mathrm{ATL}[\mathit{iR}] \subsetneq \mathrm{ATL}[\mathit{IR}]$.

3. $\mathrm{ATL}[\mathit{Ir}] = \mathrm{ATL}[\mathit{IR}]$

4. $\mathrm{ATL}^*_{\mathsf{Ir}} \subsetneq \mathrm{ATL}^*_{\mathsf{IR}}$.

5. $\mathrm{ATL}[\mathit{ir}] \subsetneq \mathrm{ATL}[\mathit{iR}]$.

Summary:

$$\mathrm{ATL}[\mathit{ir}] \overset{5}{\subsetneq} \mathrm{ATL}[\mathit{iR}] \overset{2}{\subsetneq} \mathrm{ATL}[\mathit{IR}] \; (\overset{3}{=} \mathrm{ATL}[\mathit{Ir}]) \overset{4}{\subsetneq} \mathrm{ATL}^*_{\mathsf{IR}}.$$

(Addendum) Detectives and fugitives puzzles

A fugitive is trying to run away from a detective, who is trying to catch the fugitive. There are N caves arranged in a line and the fugitive is hiding in one of them. Every midnight the fugitive must move from the current cave to one of the neighbouring caves. Every day the detective inspects one of the caves, of his choice. The detective can only catch the fugitive if he is in the inspected cave.

1. Model the scenario for a given N with a CGM $\mathcal{M}_N$.

   (Hint: assume that the fugitive and the detective are moving from one cave to another simultaneously.)

2. Does the detective have a strategy to eventually catch the fugitive, no matter where he is initially, if: $N = 4$? $N = 5$? $N > 5$?

   In each case: if not, why? If yes, what is the least number of moves within which the strategy is guaranteed to succeed?

3. Same questions, if every time the fugitive has the choice to either remain in the same cave or move to a neighbouring cave.

4. Same questions, if the caves are not arranged linearly, but are at the vertices of any neighbourhood graph.

5. Same questions, if there are two detectives, working as a team.

Epistemic extensions of ATL

The Coalitional Multi-agent Epistemic Logic ATEL

ATEL: proposed by van der Hoek and Wooldridge (2002) as a fusion of ATL and the multi-agent epistemic logic MAEL.

ATEL formulae, where A is any set of agents:

$$\varphi := p \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\langle A\rangle\rangle \mathcal{X} \varphi \mid \langle\langle A\rangle\rangle \mathcal{G} \varphi \mid \langle\langle A\rangle\rangle \varphi \,\mathcal{U}\, \psi \mid$$

$$\mathrm{K}_{i}\varphi \mid \mathrm{K}_{A}\varphi \mid \mathrm{C}_{A}\varphi \mid \mathrm{D}_{A}\varphi$$

All other propositional connectives are definable as usual and $\mathcal{F}\varphi := \top \,\mathcal{U}\, \varphi$.

Some examples:

- $\langle\langle 1\rangle\rangle \mathcal{X} \varphi \rightarrow \mathrm{K}_1 \langle\langle 1\rangle\rangle \mathcal{X} \varphi$;
- $\langle\langle 1, 2\rangle\rangle \mathcal{G} \varphi \wedge \neg \mathrm{C}_{\{1,2\}} \langle\langle 1, 2\rangle\rangle \mathcal{G} \varphi$;
- $\mathrm{D}_{\{1,2\}} \varphi \rightarrow \langle\langle 1, 2\rangle\rangle \mathcal{F} \mathrm{C}_{\{1,2\}} \varphi$;
- $\langle\langle 1, 2\rangle\rangle \mathcal{F} \varphi \rightarrow \neg \mathrm{K}_3 \neg \langle\langle 1, 2\rangle\rangle \mathcal{F} \varphi \wedge \mathrm{K}_{\{1,2\}} \neg \langle\langle 3\rangle\rangle \mathcal{G} \neg \varphi$;

Knowledge and strategic abilities of agents and coalitions

- "*Every agent in the coalition* A *knows that the coalition has a joint strategy to achieve the Goal.*"

$$K_{\mathsf{A}} \langle\langle \mathsf{A} \rangle\rangle \mathcal{F}\, \mathrm{Goal}$$

- "*It is a common knowledge in the coalition* A *that it has a collective strategy to maintain safety until reaching a winning state.*"

$$C_{\mathsf{A}} \langle\langle \mathsf{A} \rangle\rangle\, \mathrm{Safe}\, \mathcal{U}\, \mathrm{Win}$$

- Claiming existence of solution to the "Russian cards" problem:
'*If the card of player* **c** *is a distributed knowledge of* **a** *and* **b**, *then they have a strategy to make it a common knowledge between them, as well as that* **c** *has not learned any of the cards of* **a** *and* **b**'

$$\mathrm{D}_{\{1,2\}}\mathrm{Card}(\mathbf{c}) \rightarrow \langle\langle \{\mathbf{a}, \mathbf{b}\} \rangle\rangle \mathcal{F} \mathrm{C}_{\{\mathbf{a},\mathbf{b}\}}(\mathrm{Card}(\mathbf{c}) \wedge \neg K_{\mathbf{c}}(\ldots))$$

However, ATEL cannot express knowledge and coordination of strategies, e.g.:
"The agent **a** knows a strategy to achieve Goal" or,

"Every agent in the coalition A knows a joint strategy to achieve Goal, and the coalition can coordinate on that strategy", etc.

ATEL: formal semantics

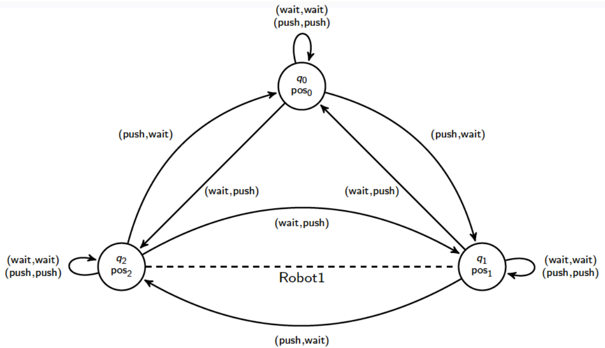
Semantic structures for ATEL: concurrent epistemic game structures

$$\langle \mathbb{A}, S, \{\sim_{\mathsf{i}}\}_{\mathsf{i}\in\mathbb{A}}, \mathsf{Act}, \mathsf{act}, \mathsf{out}, \mathsf{AP}, \mathsf{L}\rangle$$

combining concurrent game models (CGM) and multi-agent epistemic models.

The semantics for ATEL combines the CGM-based semantics for ATL and the Kripke semantics for multi-agent epistemic logics.

Deciding the truth of ATEL formulae: some examples



$\mathcal{M}, q_1 \overset{?}{\models} \mathrm{K}_1 \langle\langle 1,2 \rangle\rangle \mathcal{X}\, \mathsf{pos}_1$ Yes. $\mathcal{M}, q_1 \overset{?}{\models} \mathrm{K}_1 \langle\langle 1,2 \rangle\rangle \mathcal{X}\, \mathsf{pos}_0$ No.

$\mathcal{M}, q_1 \overset{?}{\models} \mathrm{D}_{\{1,2\}} \langle\langle 1,2 \rangle\rangle ((\neg \mathsf{pos}_2)\ \mathcal{U}\ \mathsf{pos}_0)$ Yes?

Some variations and extensions of ATL: references

- Epistemic extension of ATL:

W. van der Hoek and M.J.W. Wooldridge,
Cooperation, Knowledge, and Time: Alternating-time Temporal Epistemic Logic and its Applications, Studia Logica, 2003.

- ATL with incomplete information and memory:

P.Y. Schobbens, Alternating-time logic with imperfect recall,
Electronic Notes in Theoretical Computer Science, 2004.

- ATL with explicit knowledge of strategies:

W. Jamroga and W. van der Hoek,
Agents that Know how to Play, Fundamenta Informaticae, 2004.

- Constructive knowledge of strategies:

Thomas Ågotnes and Wojciech Jamroga,
Constructive Knowledge: What Agents Can Achieve under Imperfect Information,
Journal of Applied Non-Classical Logics, 2007.

Some conceptual deficiencies of ATEL

The semantics for ATEL is quite abstract and rather questionable. It fails to address adequately a number of important issues, such as:

- The agents' knowledge: *dynamic* or *static*?

  The static meaning of $K_{\mathbf{a}}\varphi$ is:
  '*The agent a knows that φ*'.

  The dynamic meaning is:
  '*As far as the agent a currently knows, φ is true*'.

- ATEL in its original form formalizes multi-agent systems with *static knowledge*, where agents neither learn, nor forget.

  Thus, the *dynamics of knowledge* is not taken into account.

- *The interaction between knowledge and abilities* for agents and coalitions is not reflected in the ATEL semantics, either.

The road ahead

The general challenge: *to develop a more involved and flexible semantics, capturing the dynamics of the interaction between knowledge and strategic abilities under imperfect information*.

For a proposal outline in that direction, see (the last part of)

Valentin Goranko and Eric Pacuit,
Temporal Aspects of the Dynamics of Knowledge,
in: Baltag, Alexandru, Smets, Sonja (Eds.), Johan van Benthem on Logic and Information Dynamics, Outstanding Contributions to Logic, Vol. 5, Springer, 2014, pp. 235-266.

An application to multi-agent planning

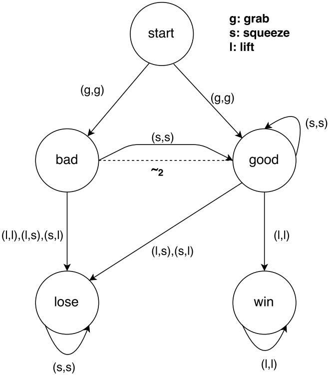
Multi-agent planning is about design of a joint strategy for the entire team of agents (possibly acting against environment) to achieve a common objective, typically reachability.

This task is reduced to model checking an ATL formula of the type $\langle\langle \mathbb{A} \rangle\rangle \mathcal{F}\, \mathrm{Goal}$. Since model checking is done constructively, proving the truth of such formula also constructs a witnessing strategy.

In the case of complete and perfect information, this model-checking task is solved easily, essentially by backward reachability analysis.

In the case of incomplete or imperfect information, positional strategies no longer suffice and the model-checking problem becomes generally undecidable (but still semi-decidable, as finite-memory strategies suffice).

Applications to multi-agent planning: example



A container with very dangerous acid is on the floor in a chemical plant. Robots 1 and 2 are to lift it together and remove it. Both are equipped with a sensor that can detect if they are touching the container. Due to humidity, the container may be slippery. The robots can squeeze to improve their grip. Robot 1 has a grip sensor that can detect whether they (both) have a good enough grip. The robots have synchronised clocks.

The task is to synthesise a joint uniform strategy that guarantees that the container is lifted without falling. The problem is that only Robot 1 has perfect information, while Robot 2 cannot distinguish between states good and bad.

There is no memoryless uniform joint strategy for this problem.

However, there is a memory-based one. How to synthesise it?

An iterated subset construction can be applied.

Recent work:

D.Gurov, V. Goranko, E. Lundberg: Knowledge-Based Strategies for Multi-Agent Teams Playing Against Nature, *Artificial Intelligence*, 2022.

Online published version: `https://www.sciencedirect.com/science/article/pii/S0004370222000686`

Lecture 2: summary and concluding remarks

Knowledge plays crucial role for the abilities of the players and coalitions to guarantee achievement of their objectives.

The standard ATL framework does not take that into account.

Incomplete knowledge is modeled in terms of players' uncertainties.

ATL can be adapted to deal with incomplete information semantically, by restricting the truth clauses of the strategic operators to uniform strategies.

Memory is important in ATL with incomplete information.

The incomplete information can also be expressed in the language, by adding explicit epistemic modal operators for individual and collective knowledge.

A major challenge: to model and take into account the dynamics of interaction between knowledge and strategic abilities of the agents.

END OF LECTURE 2