Knowledge and Beliefs

Prof. Ing. Federico Chesani

DISI

Department of Informatics – Science and Engineering

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Motivation

- Agents are in a world, and they might have incomplete knowledge about it
- Each agent might have a different knowledge
- Up to know, no way for an agent to reason upon its own knowledge, or the knowledge of another agent

FOL is powerful for describing knowledge, however it is missing useful shortcuts for dealing with concepts such as knowledge, belief, etc.

Moreover, agents act in a world that evolve along time. How to reason upon such evolution?

Goal of this lesson

- Modal Logic
- LTL Logic

Disclaimer and Further reading

Reading:

- AIMA, chapter 10, Section10.4

Further reading:

- https://www.sciencedirect.com/science/article/abs/pii/S1574652607030155
- https://www-sciencedirect-com.ezproxy.unibo.it/science/article/pii/S157465260703012X
- https://academic.oup.com/logcom/article-abstract/7/4/429/1080874

All freely downloadable through the unibo proxy

Knowledge and Belief

Our agent in the environment

So far, we discussed about representing knowledge

Implicitly, we adopted this point of view:

- our agent/program/decision system acts in a certain environment
- knowledge about the environment is available
- our agent take decisions/plan on the basis of the knowledge

Our agent in the environment

The viewpoint of an agent with KB has some limits

In the environment there are other agents

- Success of any decision depends on the actions taken by other agents
- These others agents might have their own knowledge…
- … knowing something about the others might help to take our own decisions?

Our agent has a KB, but it does not know it

- no distinction between knowledge, belief, perception, …
- no possibility of any meta-reasoning on its own knowledge

Propositional attitudes

Examples:

- There is a pizza in the fridge
- Federico knows that there is a pizza in the fridge
- Federico believes that there is a pizza in the fridge
- Federico wants a pizza
- Federico informs Elena there is a pizza in the fridge
- Elena knows there is a pizza in the fridge
- Federico knows that Elena knows there is a pizza in the fridge
- Federico informs Elena that he wants a pizza
- Elena knows Federico wants a pizza
- Federico knows Elena knows Federico wants a pizza

Propositional attitudes

Examples:

- Federico knows that there is a pizza in the fridge
- Federico believes that there is a pizza in the fridge
- Federico wants a pizza
- Federico informs Elena there is a pizza in the fridge
- Elena knows there is a pizza in the fridge
- Federico knows that Elena knows there is a pizza in the fridge
- Federico informs Elena that he wants a pizza
- Elena knows Federico wants a pizza
- Federico knows Elena knows Federico wants a pizza

The examples show what have been named propositional attitudes of an agent towards mental objects (its own, and of other agents). E.g.: **Believes**, **Knows**, **Wants**, **Informs**

Propositional attitudes

How to represent propositional attitudes? Using FOL predicates?

- they refer to something that is not a term
  Knows(Federico, location(pizza, fridge))
  location(pizza, fridge) is usually a sentence, not a term
- FOL predicates might support referential transparency
  Knows(Lois, CanFly(Superman))
  Superman=Clark
  Does Lois knows that Clark can fly?

Propositional attitudes behave differently from predicates

- using FOL predicates is not the best choice ("does not work")

Modal operators

The idea was to introduce new, special operators, named Modal operators, that behaves differently from logical operators.

Modal Logics

- same syntax as FOL
- augmented with modal operators
- modal operators take as input sentences (rather than regular terms)
- each modal operator takes two input: the name of an agent, and the sentence it refers to
- Semantics is extended with the notion of possible worlds
- Possible worlds are related through an accessibility relation (that might depend on the operator)
- Different operators define different logics

Modal operators and Modal Logics

Example: operator K for indicating that agent a knows P:

$$\mathbf{K}_a P$$

- Federico knows there is a pizza in the fridge
  $\mathbf{K}_{federico}location(pizza,fridge)$
- Federico knows that Elena knows that there is a pizza in the fridge
  $\mathbf{K}_{federico}\mathbf{K}_{elena}location(pizza,fridge)$

Modal Logics

Semantics of Modal Logics

The starting point: an agent in a given scenario is **not omniscient** regarding all the aspects of the current state of the world (the current world).

Thus, an agent might think:

- there is the current perception of the state of the world
- the unknown is depicted as possible worlds, with an intuitive notion of possibility or accessibility from my current world

We are not really interested on a specific notion of "accessibility": it might suffice that there is such a notion.

The idea is something like: an agent knows p if in any accessible world accessible from the current, p is true.

Semantics of Modal Logics – formally

Given a set of primitive propositions ϕ , a Kripke structure $M=(S, \boldsymbol{\pi}, K_1,\ldots,K_n)$ is defined:

- S is the set of states of the world
- π: $\phi \rightarrow 2^S$ specifies for each primitive proposition the set of states at which the proposition hold
- $K_1, \ldots, K_n$ are binary relations over S, i.e. $K_i \subseteq S \times S$, with the meaning: $(s, t) \in K_i$ if agent i consider the world t possible from s.

Don't get confused by the accessibility/possibility relation: we are not really interested in it.

Semantics of Modal Logics – an example

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, $\boldsymbol{\pi}$, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages: before Alice toss the coin, after Alice toss the coin, after Bob enters the room: $S = \{s_0, h_1, t_1, h_2, t_2\}$, where s_0 is the initial stage; h_1, t_1 capture that the coin has been tossed and landed head/tail; h_2, t_2 capture that Bob entered the room and observed the coin.

Semantics of Modal Logics – an example

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, $\boldsymbol{\pi}$, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages: before Alice toss the coin, after Alice toss the coin, after Bob enters the room: $S = \{s_0, h_1, t_1, h_2, t_2\}$

Which propositions can be true in any stage?

- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$

Semantics of Modal Logics – an example

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, $\boldsymbol{\pi}$, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages, five worlds $S = \{s_0, h_1, t_1, h_2, t_2\}$
- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$

Alice in the room: from her viewpoint, she always perfectly observes each world, and there is no "possible worlds" left out:

- $K_A = \{(s, s) \mid s \in S\}$

Semantics of Modal Logics – an example

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, **π**, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages, five worlds $S = \{s_0, h_1, t_1, h_2, t_2\}$
- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$
- $K_A = \{(s, s) \mid s \in S\}$

Bob observes everything in the first and the third stage, but is unsure in the second stage:

- $K_B = \{(s, s) \mid s \in S\} \cup \{ (h_1, t_1), (t_1, h_1) \}$

Semantics of Modal Logics – an example

M=(S, $\boldsymbol{\pi}$, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages, five worlds $S = \{s_0, h_1, t_1, h_2, t_2\}$
- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$
- $K_A = \{(s,s) \mid s \in S\}$
- $K_B = \{(s,s) \mid s \in S\} \cup \{ (h_1, t_1), (t_1, h_1) \}$

Finally, we can establish the truth of sentences like:

$$(M, s_0) \vDash K_A \neg tossed \wedge K_B K_A K_B (\neg head \wedge \neg tail)$$

$$(M, s_1) \vDash (head \vee tail) \wedge \neg K_B head \wedge \neg K_B tail \wedge K_B (K_A head \vee K_A tail)$$

Properties of the knowledge modal operator

Which are the axioms of our logic?

It's up to us to add the ones we deem as useful / necessary / important / whatever

Axiom **A0**: All instances of propositional tautologies are valid

MP Modus Ponens: if φ is valid and $\varphi \Rightarrow \psi$ is valid, then ψ is valid

What about the knowledge operator K w.r.t. the modus ponens?

(Distribution Axiom) **A1**: $\left(K_i\varphi \ \land \ K_i(\varphi \Rightarrow \psi)\right) \Rightarrow K_i\psi$

in other words, *knowledge is closed under implication*.

But… A1 assumes quite powerful agents!!!

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Properties of the knowledge modal operator

The definition so far of knowledge assumes quite powerful agents…

Notice:

- if φ is true in all the worlds of a structure M (i.e., φ is valid)…
- … then φ must be true, in particular, in all the worlds (M,s) that agent i considers as possible.
- If φ is true in any possible world, then agent i knows φ (by definition) in all the possible worlds.

(Knowledge Generalization Rule) **G**: for all structures M, if $M \vDash \varphi$, then $M \vDash K_i\varphi$

In other words, our agent knows all the tautologies. This of course holds for tautologies only… i.e., $\varphi \Rightarrow K_i\varphi$ is not valid (unless we axiomatize it).

Properties of the knowledge modal operator

Which is the relation between knowing something and the truth value of that?

(Knowledge Axiom) **A2**: If an agent knows φ , then φ is true

$$(K_i\varphi \Rightarrow \varphi)$$

This is considered to be the central property of Modal logic for knowledge. (valid only if the possibility/accessibility relation is reflexive)

What about belief? This axiom, in belief logic, does not hold. It is usually substituted by:

A2': $\neg K_i\ false$

Properties of the knowledge modal operator

What about the own knowledge of an agent?

(Positive Introspection Axiom) **A3**:

$$(K_i\varphi \Rightarrow K_iK_i\varphi)$$

(Negative Introspection Axiom) **A4**:

$$(\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi)$$

The agent might have a partial knowledge of the world (hence the possible worlds), but with these axioms we get sure that at least it has perfect knowledge about its own knowledge.

Modal logic – Axioms and rules

- **A0**: All instances of propositional tautologies are valid
- **MP** Modus Ponens: if φ is valid and $\varphi \Rightarrow \psi$ is valid, then ψ is valid
- (Distribution Axiom) **A1**: $\big(K_i\varphi \ \wedge \ K_i(\varphi \Rightarrow \psi)\big) \Rightarrow K_i\psi$
- (Knowledge Generalization Rule) **G**: for all structures M, if $M \vDash \varphi$, then $M \vDash K_i\varphi$
- (Knowledge Axiom) **A2**: If an agent knows φ , then φ is true $(K_i\varphi \Rightarrow \varphi)$
- (Positive Introspection Axiom) **A3**: $(K_i\varphi \Rightarrow K_iK_i\varphi)$
- (Negative Introspection Axiom) **A4**: $(\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi)$

Shall we take all of them?

- Do we need all of them?
- If not, which one will we keep?

Different Modal logics

- **A0**: All instances of propositional tautologies are valid
- **MP** Modus Ponens: if φ is valid and $\varphi \Rightarrow \psi$ is valid, then ψ is valid
- (Distribution Axiom) **A1**: $\left(K_i\varphi \ \wedge \ K_i(\varphi \Rightarrow \psi)\right) \Rightarrow K_i\psi$
- (Knowledge Generalization Rule) **G**: for all structures M, if $M \vDash \varphi$, then $M \vDash K_i\varphi$
- (Knowledge Axiom) **A2**: If an agent knows φ , then φ is true $(K_i\varphi \Rightarrow \varphi)$
- (Positive Introspection Axiom) **A3**: $(K_i\varphi \Rightarrow K_iK_i\varphi)$
- (Negative Introspection Axiom) **A4**: $(\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi)$

- A0, A1, MP, and G: **Logic K**
- A0–A4, MP, G: **Logic S5**
- A0–A3, MP, and G: **Logic S4**

It up to us decide:

- which operators
- which axioms
- and **mix the ingredients to achieve a logic**

Other operators?

- B, for belief logics
- E (everyone knows), for group knowledge
- CK (common knowledge), for group knowledge
- D (distribute knowledge), for group knowledge
- …

Linear-time Temporal Logic (LTL)

What about time? and dynamics of the worlds?

The world observed by an agent evolves through time…

… the framework of modal logic has been exploited to support a qualitative representation and reasoning over the evolution of worlds along the temporal dimension

- the accessibility relation now is mapped into the temporal dimension
- each world is labelled with an integer: time is then discrete

Two possible ways for the world to evolve:

- Linear-time: from each world, there in only one other accessible world
- Branching-time: from each world, many worlds can be accessed

Temporal logics

Which operators?

(remember, we can choose…)

Temporal logics

Which operators? Notice that we chose to organize time along a discrete evolution of worlds

- Something is true at the next moment in time
  $\bigcirc\varphi$
- There is a φ that is always true in the future
  $$\square\varphi$$
- There is a φ that is true sometimes in the future
  $$\diamond\varphi$$

- There exists a moment when ψ holds and φ will continuously hold from now until this moment
  $$\varphi\mathcal{U}\psi$$
- φ will continuously hold from now on unless ψ occurs, in which case φ will cease
  $\varphi\mathcal{W}\psi$

Temporal Logics – semantics

The semantics is defined again starting from a Kripke strcuture M.
However:

- the accessibility relation is mapped on a linear structure, where each world/state is simply labelled with a natural number…
- … thus indicating the state or the number is the same

A proposition is entailed when:

- $(M, i) \vDash p$ iff $\mathrm{i} \in \pi(p)$
- $(M, i) \vDash \bigcirc\varphi$ iff $(M, i + 1) \vDash \varphi$
- $(M, i) \vDash \square\varphi$ iff $(M, j) \vDash \varphi,\ \forall j \geq i$

- $(M, i) \vDash \varphi \mathcal{U} \psi$ iff $\exists k \geq i\ s.t. (M, k) \vDash \psi$ and $\forall j. i \leq j \leq k\ (M, j) \vDash \varphi$
- $(M, i) \vDash \varphi \mathcal{W} \psi$ iff either $(M, i) \vDash \varphi \mathcal{U} \psi$ or $(M, i) \vDash \square\varphi$

Temporal Logic – the example

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, **π**, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages, five worlds $S = \{s_0, h_1, t_1, h_2, t_2\}$
- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$
- $K_A = \{(s, s) \mid s \in S\}$
- $K_B = \{(s, s) \mid s \in S\} \cup \{ (h_1, t_1), (t_1, h_1) \}$

$$(M, 0) \vDash \bigcirc tossed \ \wedge K_B \bigcirc (heads \vee tail) \wedge K_B \Box (tail \Rightarrow K_A tail)$$

Temporal Logic and other operators

- Alice is in a room, and will toss a coin, the coin will land head or tail.
- Bob is in another room, he will hear the sound of the coin on the floor, will enter the room and observe if the coin is head or tail.

M=(S, $\boldsymbol{\pi}$, K_a, K_b):

- $\Phi = \{tossed, head, tail\}$
- Three stages, five worlds $S = \{s_0, h_1, t_1, h_2, t_2\}$
- $\pi(tossed) = \{h_1, t_1, h_2, t_2\}$
- $\pi(head) = \{h_1, h_2\}$
- $\pi(tail) = \{t_1, t_2\}$
- $K_A = \{(s, s) \mid s \in S\}$
- $K_B = \{(s, s) \mid s \in S\} \cup \{ (h_1, t_1), (t_1, h_1) \}$

$$(M, 0) \vDash \bigcirc tossed \ \wedge K_B \bigcirc (heads \vee tail) \wedge K_B \Box (tail \Rightarrow K_A tail)$$

Model checking

LTL (but also CTL) lend themselves nicely to the modelling of system specifications that can be considered as finite state machines.

They proved quite effective also in modelling distributed systems, where different agents can exchange messages and information.

A question arises: given a specification of a (distributed) system, is it possible to prove that some properties always/sometimes hold?

- Is it possible to prove that a certain state will never occur? (safety properties)
- Is it possible to prove that a certain state will occur, sooner or later?
- Is it possible to prove that my system will not end up in a loop? (livelock/deadlock)

Model checking

Model Checking is a discipline that investigated the possibility of proving such formal properties.

Many approaches and tools have been proposed, among them many tools adopted the LTL semantics.

Many applications fields benefitted from these techniques:

- hardware design
- safety-critical systems
- safety-critical processes and oeprations
- security protocols

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

federico.chesani@unibo.it