# Agent-Based Systems

Michael Rovatsos
mrovatso@inf.ed.ac.uk

Lecture 14 – Logics for Multiagent Systems

# Where are we?

Last time . . .

- Argumentation: a richer form of negotiation
- Logic-based negotiation: attacks, defeats
- Strengths of arguments
- Abstract argumentation systems
- (Implemented) argumentation dialogue systems

Today . . .

- **Logics for Multiagent Systems**

# Logics for multiagent systems

- Throughout computer science, logic is used to develop formal models of computation
- In multiagent systems, the predominant approach for doing this is based on **modal logics**
- These are used to model agents' mental states (but also other approaches, e.g. modelling commitments, obligations and permissions, etc)
- We will first introduce the most common model of modal logic semantics, then use it to model beliefs and knowledge

# Why modal logic?

- We are looking for a logic to describe mental states
- Consider the following statement:

  *Michael believes Kylie likes the ABS course*

- Naive attempt: use first-order logic (FOL) to express this, i.e.

$$Bel(Michael, Likes(Kylie, ABS))$$

- But this is not a syntactically correct FOL formula (terms cannot be predicates)!
- We could think of "$Likes(Kylie, ABS)$" as an object (a constant), but that's not really elegant

# Why modal logic?

- The semantic problem is even worse:
  - Kylie is a student ➡ we can accept statement $Kylie = s987654$
  - But would we conjecture that $Bel(Michael, Likes(s987654, ABS))$? After all, Michael might not know about this equality ...
  - Problem: intentional notions are **referentially opaque**, they set up opaque contexts in which FOL substitution rules don't apply
- Classical logic based on **truth functional** operators: the truth value of $p \wedge q$ is a function of the truth values of $p$ and $q$
  - Semantic value (denotation) of a formula depends only on denotations of sub-expressions
  - But "Michael believes $p$" is not truth-functional, it depends on truth value of $p$ *and* Michael's belief
  - So substitution will not preserve meaning and won't work

# Possible-worlds semantics

- Kripke's (1963) model of possible worlds: standard for modal logic semantics
- Example: a game of cards, agents cannot see each others set of cards
  - useful for agent to infer which cards are held by others
  - consider all alternative distributions of cards among all players
  - own cards (and cards on the table) eliminate certain alternatives
  - remaining possible combinations of sets of cards is a possible world
- We can describe the agents belief by the set of worlds he thinks possible **epistemic alternatives**

# Normal modal logic

- Before moving to **epistemic logic** we describe the framework of **normal modal logic** as its foundation
- Based on distinction between *necessary* and *contingent* truths
- Necessary truths are true in all possible worlds, possible truths are true in some possible worlds
- Use $\Box$ (box) and $\Diamond$ (diamond) operators to denote “necessarily” and “possibly”
- We introduce a simple propositional modal logic (like classical propositional logic extended with the two modal operators)

# Normal modal logic – Syntax

- Syntax of our language given by defining what its formulae are
- Let $Prop = \{p, q, \ldots\}$ countable set of atomic propositions
- If $p \in Prop$, $p$ is a formula
- If $\varphi$, $\psi$ are formulae, then so are

$$true \quad \neg\varphi \quad \varphi \vee \psi$$

  with the usual meaning as in ordinary propositional logic
- Other operators ($\wedge$, $\Rightarrow$) and the constant *false* can be defined as abbreviations of the above
- If $\varphi$ is a formula, then so are $\Box\varphi$ and $\Diamond\varphi$

# Normal modal logic – Semantics

- Let $W$ a set of worlds, $R \subseteq W \times W$ an **accessibility relation** describing which worlds are possible relative to other worlds
- $\langle W, R, \pi \rangle$ is a **model** for normal propositional modal logic with valuation function $\pi : W \rightarrow \wp(Prop)$
- $\pi$ specifies which atomic propositions are true in which world
- **Satisfiability** relation $\models$ between pairs $\langle M, w \rangle$ and formulae of the language used to define semantics:
  - $\langle M, w \rangle \models true$
  - $\langle M, w \rangle \models p$ iff $p \in \pi(w)$
  - $\langle M, w \rangle \models \neg\varphi$ iff $\langle M, w \rangle \not\models \varphi$
  - $\langle M, w \rangle \models \varphi \vee \psi$ iff $\langle M, w \rangle \models \varphi$ or $\langle M, w \rangle \models \psi$
  - $\langle M, w \rangle \models \Box\varphi$ iff $\forall (w, w') \in R. \langle M, w' \rangle \models \varphi$
  - $\langle M, w \rangle \models \Diamond\varphi$ iff $\exists (w, w') \in R. \langle M, w' \rangle \models \varphi$
- Modal operators are **duals** of each other: $\Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ (like $\exists/\forall$)

# Correspondence theory

- A formula is called
  - satisfiable if it is satisfied for some model/world pair
  - unsatisfiable if it is not satisfied for any model/world pair
  - true in a model if it is satisfied for every world in the model
  - valid in a class of models if it is true in every model in the class
  - valid if it is true in the class of all models
- If $\varphi$ is valid, we write $\models \varphi$ (all tautologies in propositional logic are valid)
- Two basic properties:
  - **K-axiom**: $\models \Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi)$ is a valid formula
  - **Necessitation rule**: If $\models \varphi$ then $\models \Box\varphi$
- These appear in any complete axiomatisation of normal modal logic, but turn out to be the most problematic . . .

# Correspondence theory

- A **system of logic** is a set of formulae valid in some class of models
- A member $\varphi$ of this set is called a **theorem** of the logic ($\vdash \varphi$)
- Different sets of axioms correspond to different properties of the accessibility relation $R$ (**correspondence theory**)
- Axioms are **characteristic** of a class of models if they are satisfied by all and only those models
- $K\Sigma_1 \ldots \Sigma_n$ refers to the smallest modal logic containing axioms $\Sigma_1 \ldots \Sigma_n$

## Correspondence theory

- Correspondence between properties of $R$ and axioms:

| Name | Axiom | Property of $R$ | Characterisation |
|---|---|---|---|
| T | $\Box\varphi \Rightarrow \varphi$ | Reflexive | $\forall w\ .(w, w) \in R$ |
| D | $\Box\varphi \Rightarrow \Diamond\varphi$ | Serial | $\forall w\ \exists w'\ .(w, w') \in R$ |
| 4 | $\Box\varphi \Rightarrow \Box\Box\varphi$ | Transitive | $\forall w, w', w''\ .(w, w') \in R \wedge (w', w'') \in R \Rightarrow (w, w'') \in R$ |
| 5 | $\Diamond\varphi \Rightarrow \Box\Diamond\varphi$ | Euclidean | $\forall w, w', w''\ .(w, w') \in R \wedge (w, w'') \in R \Rightarrow (w', w'') \in R$ |

- Interestingly, instead of $2^4 = 16$ systems of logic there are only 11 because some are equivalent (contain the same theorems)
- Some abbreviations often used: $KT$ is called T, $KT4$ is called S4, $KD45$ is weak-S5, $KT5$ called S5

# Normal modal logics as epistemic logics

- Looking at single agent knowledge, we can assume that the agent knows something if it is true in all accessible possible worlds
- We can use $\Box\varphi$ to denote "it is known that $\varphi$"
- In the case of several agents, models have to be extended to structures

  $$\langle W, R_1, \ldots, R_n, \pi\rangle$$

  where $R_i$ accessibility relation of $i$
- The single modal operator $\Box$ is replaced by unary modal operators $K_i$, one for each agents
- We replace rule for "$\Box$" by

  $$\langle M, w\rangle \models K_i\varphi \text{ iff } \forall(w, w') \in R_i.\langle M, w'\rangle \models \varphi$$

- The systems of logic above can be extended accordingly (e.g. S5 becomes $S5_n$)

## Normal modal logics as epistemic logics

- How well-suited are the properties of normal modal logic for describing knowledge and belief?
- Necessitation rule means that agents know all valid formulae (amongst others the tautologies of propositional logic)
- So agents always have an infinite amount of knowledge ➡ counterintuitive
- K-axiom causes a similar problem
    - Suppose $\varphi$ is logical consequence of $\{\varphi_1, \ldots \varphi_n\}$
    - $\varphi$ is true in every world in which $\varphi_1, \ldots \varphi_n$ are
    - Therefore $\varphi_1 \wedge \cdots \wedge \varphi_n \Rightarrow \varphi$ is valid
    - By necessitation, this rule must be believed
- By the K-axiom, the agent's knowledge is closed under logical consequence (if agent believes premises, it believes consequence)
- Agents know everything they might be able to infer!

# Logical omniscience

- **Logical omniscience** problem: knowing all valid formulae and knowledge/belief being closed under logical consequence
- One problem concerns consistency: human reasoners often have beliefs $\varphi$ and $\psi$ with $\varphi \vdash \neg\psi$ without being aware of inconsistency
- Ideal reasoners would believe *every formula* of the logic in this case
- This is because the consequential closure of “false” is the set of all formulae
- More reasonable to require **non-contradictory beliefs**, i.e. that $\varphi$ and $\neg\varphi$ are not believed at the same time

# Logical omniscience

- Second problem concerns logical equivalence
- Example: Assume we believe the following propositions
  1. Hamlet's favourite colour is black
  2. Hamlet's favourite colour is black *and* every planar map can be four coloured
- 2. will be believed if and only if 1. is believed, i.e. they are logically equivalent
- But equivalent propositions should not be equivalent as beliefs!
- Yet this is what possible-worlds semantics implies
- It has been argued that propositions are thus too coarse grained to serve as beliefs in this way

## Axioms for knowledge and belief

- How appropriate are the axioms D, T, 4, and 5 for logics of knowledge and belief?
- Axiom D requires that beliefs are not contradictory (reasonable):

$$K_i\varphi \Rightarrow \neg K_i \neg\varphi$$

- Axiom T often called **knowledge axiom**, requires that everything that is known is true
- This can be used to distinguish knowledge from belief such that "*i* knows $\varphi$ if *i* believes $\varphi$ and $\varphi$ is true"

# Axi oms for knowledge and belief

- Defining knowledge in this way it satisfies T
- Axioms 4/5 is called **positive/negative introspection** meaning that an agent knows what it knows/doesn't know
- Negative introspection considered more demanding than 4
- Usually, S5 is chosen as a logic of knowledge and KD45 as a knowledge of belief

## Common and distributed knowledge

- One would also like to model **common knowledge**, i.e. the things everyone knows, things everyone knows that everyone knows, etc.
- Introduce an operator for "everyone knows $\varphi$" as an abbreviation

$$E\varphi := K_1\varphi \wedge \cdots \wedge K_n\varphi$$

- But this is not enough, it doesn't describe that everyone is aware that everyone knows $\varphi$ (and so on)
- Define another operator $C$ for "it is commonly known that $\varphi$"
  - Let $E^1\varphi := E\varphi$ and $E^{k+1}\varphi := E(E^k_\varphi)$
  - Define $C\varphi := E\varphi \wedge E^2\varphi \wedge \cdots$
- Infinite conjunction is quite a strong requirement, does common knowledge in this sense occur in practice?

# Example

- Coordinated attack problem: two divisions of an army are camped on two hilltops waiting to attack enemy in the valley
- They can only attack successfully if they both attack at the same time
- Divisions can only communicate through messengers, communication takes time and may fail
- Even if messenger reaches other camp (e.g. with message “attack at dawn”) generals can never be sure the message was received
- Awaiting confirmation does not solve problem, confirming party will never know whether other party received confirmation
- It turns out that no amount of communication is sufficient to bring about common knowledge

## Common and distributed knowledge

- Another associated problem: distributed, implicit knowledge
- Assume an agent could read all other agents' minds ➡ this agent could have more knowledge than any other individual agent
- Example: one agent knows $\varphi$, the other (only) $\varphi \Rightarrow \psi$, omniscient observer could infer $\psi$
- Distributed knowledge operator $D$ can be introduced:

$$\langle M, w \rangle \models D\varphi \Leftrightarrow \forall (w, w') \in (R_1 \cap \cdots \cap R_n) \,.\langle M, w' \rangle \models \varphi$$

- Note that use of intersection rather than union actually *increases* knowledge
- These operators form a hierarchy:

$$C\varphi \Rightarrow E^k\varphi \Rightarrow \cdots \Rightarrow E\varphi \Rightarrow K_i\varphi \Rightarrow D\varphi$$

# Critique

- Are these complex logical models of any practical use?
- Highly valuable for system specification
- . . . but not directly implementable
- Inference intractable in most of these complex logics
- . . . we can only use them “externally”
- This doesn’t tell us anything about reasoning capabilities of agents themselves

# Summary

- Logics for multiagent systems
- Logical modelling of mental states
- Modal logic as a popular method for doing that
- Possible-world semantics, correspondence theory
- Normal modal logics as epistemic logics
- Logical omniscience problems, critique
- Epistemic logic: common knowledge, distributed knowledge
- Next time: **Summary and Concluding Remarks**