Logics for Agency and Multi-Agent Systems

Introduction to Modal Logic

Jan Broersen
Andreas Herzig
Nicolas Troquard

ESSLLI’07, Dublin, Monday, August 13, 2007

Modeling MAS

• model the agents’ mental attitudes
  – beliefs: epistemic logic, doxastic logic
  – goals, preferences: logic of desire, intention, preference
  – hopes, fears, regrets, …

• model change
  – events: dynamic logic PDL, dynamic epistemic logics DEL
  – abilities: coalition logic CL, alternating-time temporal logic ATL
  – actions, agency: STIT theory, ‘bringing it about’ logics
  – belief updates, belief revisions: KM theory, AGM theory

• model interaction and communication
  – group ability, group beliefs: ATL, common knowledge
  – commitments, norms: deontic logic
  – speech acts: speech act theory
  – protocols: dialogue games, …

Overview of today’s introduction

1. What is logic, and what do we need it for in multi-agent systems?
2. The basics of modal logic.
   1. Basic notions of logic
   2. Normal multi-modal logics
   3. Expressivity
   4. Soundness and completeness
   5. Combining logics: products
   6. Decidability and complexity
   7. Weak modal logics
3. Applications of modal logic to some reasoning domains relevant for multi-agent systems

What is logic?

• Logic is the theory of valid principles of inference, i.e. the theory that says what can be derived from information we assume to be true.

• Thus, logic says nothing about how to establish truth. Logic is only concerned with valid relations between possible truths.

• Example: the logic principle \( \neg (\phi \land \neg \phi) \) says nothing about whether \( p \) and \( \neg p \) are true or not, but only that they cannot be true at the same time.
Why is Logic relevant for MAS? (reason 1)
1. Intelligent agents need to represent their environment in order to be able to act intelligently ⇒ knowledge representation
   • Assumption (McCarthy 1958, who also suggested the term ‘Artificial Intelligence’): an efficient and intuitive representation can be found using logic.
   • The idea: agents ‘only’ need to store basic facts and assumptions, and can derive the rest through logic inference.
   • For knowledge representation, we typically need logics of common sense, concept description, conditionality, decisions, etc.

Why is Logic relevant for MAS? (reason 2)
2. In order to build agents we need to have a theory about how we want them to function ⇒ agent / system specification
   • Logic as a way of representing our ‘engineering’ knowledge.
   • If agents need to represent knowledge about other agents, we can use these logics also for knowledge representation.
   • For agent specification, we typically need logics of action, time, knowledge, belief, norms, desires, intentions, etc. ⇒ C&L, BDI and KARO

What is ‘a’ logic, formally?
• Formally a logic is a subset of some language (Later: how can this be regarded a formalization of inference patterns?)
  • Propositional logic is a subset of the language φ, ψ, ..., ::= p | ¬φ | φ ∧ ψ
  • Other logical connectives are introduced as abbreviations
  • The subset (i.e., logic) captures the logical invariants of the reasoning domain
  • The subset (i.e., logic) can be described in at least two ways:
    – Semantically, using model theory
    – Syntactically, using axiomatic systems

Basic logic notions: axiomatics
• Axiom schema’s, like φ ∨ ¬φ describe basic theorems of the logic
• In a schema, we can perform uniform substitutions / the φ are meta variables over the language
• Rules, like modus ponens, describe what theorems can be derived from which other theorems
• A formula / schema whose negation is not derivable is called consistent!
  Note: in the early days, (modal) logic was an axiomatic enterprise. Model theory had to be developed to prove logics are consistent!

Basic logic notions: semantics
• Propositional Logics (PL): a model is an assignment of truth values to atomic propositions
• Formulas are true on a model according to truth-functionality of the connectives.
• A formula is satisfiable if there is a model for it (a model that makes it true)
• A formula is valid if it is true for all possible models

Logical consequence / deduction
• If we want to use logic for knowledge representation we need to know how to derive something!
• How can a subset of some language be a formalization of reasoning patterns?
• Does the logical connective ‘→’ define logical consequence? No!
• Deduction for PL: ψ follows from φ, notation φ ⊢ ψ, if addition of φ to the axioms enables derivation of ψ
• Logical consequence for PL: ψ follows from φ, notation φ ⊨ ψ, if all models for φ are models for ψ
What is Modal Logic?

The language

Given a set of proposition symbols $P$, and $p$ ranging over $P$, the well-formed formulas $\phi$, $\psi$, ... of the language $L_M$ are generated by the 'BNF':

$\phi, \psi, \ldots := p \mid \neg \phi \mid \phi \land \psi \mid \phi \lor \psi \mid \Box \phi \mid \Diamond \phi$

Intuitively: $\Box \phi$ is 'necessarily $\phi$'
$\Diamond \phi$ is 'possibly $\phi$'

Multi-modal due to the indexes 'i'

Axiomatic answer

At least the axioms:

– All (or enough) propositional axioms are in $K$
– $\Box \phi \land \Box (\phi \rightarrow \psi) \rightarrow \Box \psi$ is in $K$

At least the rules:

– necessitation: if $\phi$ in $K$, derive that $\Box \phi$ in $K$.
– modus ponens: if $\phi$ in $K$, and $\phi \rightarrow \psi$ in $K$, derive that $\psi$ in $K$.

• Modal logics that subsume these, are called normal
• We denote that some formula $\phi$ is derivable in $K$ by: $\vdash_K \phi$

Semantic answer

• Modal logic: the theory of valid principles of inference for domains,
  – that fit well with an abstract representation consisting of a binary relation over ‘worlds’.
  – where ‘truth’ is defined relative to, a ‘current’ world

In ‘Modal Logic’ by Blackburn, de Rijke, Venema (2001) this is called ‘the local perspective’

Typical application domains of modal logics

• Time:
  – binary relation between moments
  – truth is relative to what counts as ‘the present’

• Action:
  – binary relation between (system, agent) states
  – truth is relative to what counts as ‘the current state of affairs’

• Knowledge:
  – binary relation between epistemically possible worlds
  – truth is relative to what counts as ‘the current epistemic state’

The model theory of modal logic: Kripke Models

A Kripke frame consists of a
(possibly infinite) set $W$ of possible worlds
• a relation $R$ (i.e. a possibly infinite set of tuples $(s,s')$) over $W \times W$

A model is a frame with in addition a
• valuation $V$: ATM $\rightarrow$ $2^W$ of atomic propositions

For any formula $\phi$ of a modal language $L_M$, truth of $\phi$ in a world $w \in W$ is denoted $M,w \models \phi$

Semantics

• Propositional logic connectives obey the standard truth conditions
  • The truth condition for $M,w \models \Box \phi$ is:
    for all $w'$ such that $(w,w') \in R$, it holds that $M,w' \models \phi$
  • The truth condition for $M,w \models \Diamond \phi$ is:
    there is a $w'$ such that $(w,w') \in R$, and $M,w' \models \phi$

Intuitively: $\Box \phi$ is ‘necessarily $\phi$’
$\Diamond \phi$ is ‘possibly $\phi$’
Semantics

- A formula is true on a Kripke model if and only if it is true in all states of that model.
- A formula is valid on a Kripke frame if and only if it is true on all models corresponding to that frame.
- A formula is valid (notation: $\models \phi$) with respect to a class of frames if and only if it is valid on all frames in the class.
- The set of validities of a class of frames is called ‘the logic’ of the class of frames.
- K is the logic of the class of all Kripke frames.

Some (non-)validities of basic modal logic

All propositional validities (modal logic subsumes propositional logic)

$\models \Box \psi \land \Box \phi \rightarrow \Box (\psi \land \phi)$

$\models \Box \phi \land \Box \psi \rightarrow \Box (\phi \land \psi)$

$\models \Box (\phi \land \psi) \rightarrow \Box \phi \land \Box \psi$

$\models \Box \phi \lor \Box \psi \rightarrow \Box (\phi \lor \psi)$

$\models \Box (\phi \lor \psi) \rightarrow \Box \phi \lor \Box \psi$

$\models \Box (\phi \rightarrow \psi) \rightarrow \Box (\Box \phi \rightarrow \Box \psi)$

$\models \Box \phi \land \Box (\phi \rightarrow \psi) \rightarrow \Box \phi \land \Box \psi$

Logics over particular classes of frames

Now we are going to consider logics over subclasses of the general class of all Kripke models
Rightarrow logics stronger than K.

We strengthen K by considering the logics over models generated by frames with a certain (first- or second-order) property.

In stronger logics there are more validities, formulas have more logical consequences.

Extreme case: the logic of the empty Kripke frame is the strongest possible logic, where everything is a validity, and everything is derivable.

Names for classes of models

- $T$ is the class of all reflexive Kripke models.
- $S4$ is the class of all reflexive-transitive Kripke models.
- $S5$ is the class of all Kripke models with accessibility relations that are equivalence relations.
- $KD45$ is the class of all Kripke models with serial, transitive and euclidean accessibility relations.

Well-known possible first-order properties of frames

- $F = (S, R)$ is reflexive if $\forall s \in S: (s, s) \in R$.
- $F = (S, R)$ is transitive if $\forall s, t, u \in S: (s, t) \in R \land (t, u) \in R \Rightarrow (s, u) \in R$.
- $F = (S, R)$ is symmetrical if $\forall s, t \in S: (s, t) \in R \Rightarrow (t, s) \in R$.
- $F = (S, R)$ is euclidean if $\forall s, t, u \in S: (s, t) \in R \land (s, u) \in R \Rightarrow (t, u) \in R$.
- $F = (S, R)$ is serial if $\forall S \in S, \exists t \in S: (s, t) \in R$.
- $F = (S, R)$ is an equivalence relation if R is reflexive, transitive and symmetrical.

Extra validities

- Over the model class $T$ we have e.g., the extra validity: $\Box \phi \rightarrow \phi$
- Over $S4$ we have extra e.g.: $\Box \phi \rightarrow \phi$ and $\Box \phi \rightarrow \Box \Box \phi$
- Over $S5$ we have extra e.g.: $\Box \phi \rightarrow \phi$ and $\Box \phi \rightarrow \Box \Box \phi$ and $\neg \Box \phi \rightarrow \Box \neg \Box \phi$
- Over $KD45$ we have extra e.g.: $\neg \Box \bot$ and $\Box \phi \rightarrow \Box \Box \phi$ and $\neg \Box \phi \rightarrow \Box \neg \Box \phi$.
Dependencies of frame properties

- $R$ is symmetrical and transitive $\Rightarrow$ $R$ is euclidean.
- $R$ is reflexive $\Rightarrow$ $R$ is serial.
- $R$ is symmetrical, transitive and serial $\Leftrightarrow$ $R$ is reflexive and euclidean $\Leftrightarrow$ $R$ is an equivalence relation.

These give us relations between the logics over the mentioned classes: inclusions of classes give inclusions of logics.

Two semantic modal consequence relations

Two modal consequence relations, a local and a global one:

- $\phi \Vdash_L \psi$ if $\forall M,w: M,w \Vdash \phi$ implies $M,w \Vdash \psi$.
- $\phi \Vdash_G \psi$ if $\forall M: M \Vdash \phi$ implies $M \Vdash \psi$.

and more are possible...

- we have: $\phi \Vdash_L \psi \Rightarrow \phi \Vdash_G \psi$.
- the global one is thus ‘stronger’
- the local one is most used or assumed

Differences between $\Vdash_L$ and $\Vdash_M$

- $\phi \Vdash_L \psi$
- $\phi \Vdash_G \psi$

- For instance, we have: $\phi \nvdash_L \Box \phi$, but we do have: $\phi \vdash_G \Box \phi$.
- And we have: $\phi \Vdash_L \psi$ implies $\Vdash \phi \rightarrow \psi$ (the deduction property, Hilbert 1930)
  but: $\phi \Vdash_G \psi$ does not imply $\Vdash \phi \rightarrow \psi$.

Modal logic deduction

- Again, we might want to say: $\psi$ follows from $\phi$ if addition of $\phi$ to the axioms enables derivation of $\psi$.
- But, this would correspond to global consequence, since, for instance $p \nvdash \Box p$, by the necessitation rule.
- For local consequence, which is most used, we have to constrain application of the necessitation rule to theorems only.

Expressivity

Expressivity is about what semantic entities a language is able to distinguish.

- On the level of worlds: bisimulation invariance: modal formulas cannot distinguish between bisimilar worlds.
- On the level of frames: correspondence theory: what properties of frames can be characterized by a modal logic?

Bisimulation invariance

A world $w_i$ of model $M_i$ is bisimilar to a world $w_j$ in a model $M_j$ if:

1. $V_i(w_i) = V_j(w_j)$
2. $\forall w'_1 \in W_i: w_i R_i w'_1 \text{ implies } \exists w'_2 \in W_j: w'_2 R_j w'_2$ and $w'_1$ and $w'_2$ are bisimilar
3. $\forall w'_2 \in W_j: w_j R_j w'_2 \text{ implies } \exists w'_1 \in W_i: w'_1 R_i w'_1$ and $w'_2$ and $w'_1$ are bisimilar

Intuitively, worlds are bisimilar if from these worlds, looking towards the accessible worlds, we see the same possible continuations, modulo trivial copies.
Correspondence theory

We already encountered: which modal validities are generated by a given frame class? i.e., which validities \( \phi \) hold on all frames with (first order) property \( C \)?

But, what about the other way round? Which are the frame classes that give rise to a given validity schema? i.e., which classes of frames (with first order property \( C \)) 'satisfy' the validities \( \phi \)?

If we have both ways, then the schema \( \phi \) corresponds to the FO property \( C \) on frames.

Correspondences

Example 1: the validities \( \Box \Diamond \phi \to \Diamond \phi \) correspond to the class of transitive frames \( (C = \forall s.t. sRt \land tRu \to sRu) \)

Example 2: the validities \( \Box \phi \to \Box \phi \) correspond to the class of reflexive frames \( (C = \forall w: wRw) \)

Example 3: the validities \( \Box \phi \to \Diamond \Box \phi \) correspond to the class of symmetric frames \( (C = \forall s,t: sRt \to tRs) \)

Note that these formulas say nothing about similar correspondences with classes of models: it is easy to find a model that obeys \( \phi \to \Diamond \phi \), and that is not reflexive.

Some facts on correspondences

• Not all validity schema's correspond to a first order property of frames.
  Examples:
  (1) Löb's axiom \( \Diamond (\Box \phi \to \phi) \to \Box \phi \)
  Corresponds with 'R is transitive and well-founded'
  (2) Mc Kinsey's axiom: \( \Box \Diamond \phi \to \Diamond \Box \phi \)
  • Not all first order properties of frames correspond to a modal validity schema.
  Example: irreflexivity \( C = \forall w: \neg (wRw) \)

Soundness and Completeness

Soundness

• Soundness: if \( \models \phi \) then \( \vdash \phi \)

How do we prove soundness?

Prove that:
1. axioms are validities
2. rules preserve validity

Which is often simple.

Weak completeness for modal logics

• Weak completeness: if \( \models \phi \) then \( \vdash \phi \)

Equivalent formulations of weak completeness:

1. if \( \models \phi \) then \( \vdash \phi \)
2. if \( \neg \phi \) consistent then \( \exists M,w : M,w \notmodels \phi \)
   (recall: consistency of \( \phi \) is defined as \( \notmodels \neg \phi \))
3. if \( \neg \phi \) consistent then \( \exists M,w : M,w \models \neg \phi \)
4. if \( \phi \) consistent then \( \exists M,w : M,w \models \phi \)

So, we need to prove that for every formula consistency implies satisfiability.
Completeness of logical consequence

- Weak completeness is only with respect to validities. What about logical consequence?
- If \( \Phi \models L \phi \) then \( \Phi \vdash \phi \) ?
- For the weakest notion of logical consequence (\( \models L \)), we have: \( L \phi \rightarrow L \psi \) iff \( \Phi \models L \psi \).
- It follows that for weak consequence and for finite sets \( \Phi \) we have that weak completeness (\( \models \phi \) iff \( \Phi \models \phi \)) implies: if \( \Phi \models L \phi \) then \( \Phi \vdash \phi \).
- For the stronger notions of consequence we need stronger deductive systems!

Proving strong completeness

How do we prove strong completeness?

- Build maximal consistent sets of formulas, using the language structure and the axiomatization.
- The MCSs become worlds of a ‘canonical’ Kripke model.
- Build an appropriate relational structure such that any consistent set of formulas is satisfied in some world of the canonical model.

Maximal consistent sets

A set of formulas \( \Phi \) is maximally S-consistent (is an ‘MCS’) if:
1. \( \Phi \) is S-consistent
2. no set \( \Phi \cup \{ \phi \} \) with \( \phi \notin \Phi \) is S-consistent

Property: all elements of \( S \) (i.e. all theorems) belong to any maximal consistent set.
In particular, all theorems of \( K \) belong to all maximal \( K \)-consistent sets.

Lindenbaum’s lemma

- Every S-consistent set \( \Phi \) can be extended to a maximal S-consistent set \( \Phi_{\omega} \).

Proof Lindenbaum’s lemma

Enumerate the formulas of the modal language: \( \phi_1, \phi_2, \ldots \).
Define
- \( \Phi_0 = \Phi \)
- \( \Phi_{m+1} = \Phi_m \cup \{ \phi \} \) is inconsistent, \( \Phi_{m+1} \cup \{ \phi \} \), otherwise
- \( \Phi_{\omega} = \bigcup_{n \geq 0} \Phi_n \)

Lindenbaum’s lemma (proof)

\( \Phi_{\omega} \) is consistent, because it is build by adding consistent formulas one at the time \( \Phi_{\omega} \) is maximally consistent.

Suppose not.
From definition MCS: there is a formula \( \phi_m \) such that \( \Phi_m \notin \Phi_{\omega} \) and \( \Phi_{\omega} \cup \{ \phi_m \} \) is consistent.
From definition \( \Phi_{\omega} \): \( \Phi_m \cup \{ \phi_m \} \) is inconsistent.
From definition \( \Phi_{\omega} \): \( \Phi_m \subset \Phi_{\omega} \)
From definition consistency: \( \Phi_{\omega} \) is inconsistent \( \Rightarrow \) contradiction with (1)
A canonical model

MSCs are taken as worlds $S^c$, both atomic valuations $\pi^c$ and the relation $R^c$ are defined in terms of formula membership over worlds $S^c$: 

$M^c = (S^c, \pi^c, R^c)$ with: 
$S^c = (s_\phi \mid \phi$ is an MCS) 
$\pi^c \subseteq \forall \phi (p) \iff p \in \Theta$ 
$R^c = \{(s_\pi, s_\psi) \mid \forall \phi (\psi) \in \Theta$ implies $\phi \in \Psi\}$

Note: formulas have become semantic entities (worlds)!

The truth lemma

$M^c, s_\phi \models \phi$ iff $\phi \in $Φ

This ensures that every (infinite) consistent set is satisfiable.

**Proof truth lemma**

Easy direction: $\Leftarrow$

By straightforward induction over formula structure.

More difficult direction: $\Rightarrow$

If $M^c, s_\phi \models \phi$ then $\phi \in $Φ

We only proof the case where $\phi$ is non-modal.

**Assume:** $M^c, s_\pi \models \psi$

**Def:** $\Theta \subseteq \forall \phi (\psi) \iff \psi \in \Theta$

Thus, we get $\psi \in \Theta$ by first selecting all formulas starting with a box, and then removing these initial boxes.

**Claim 1:** $\forall \phi (\neg \psi)$ is inconsistent

**Proof Claim 1**

Suppose not, i.e., $\forall \phi (\neg \psi)$ is consistent

Then by Lindenbaum: there is a MCS $\Psi$ such that $\forall \phi (\neg \psi) \subseteq \Psi$

Then, since $\forall \phi (\neg \psi) \subseteq \Psi$, we have: $\forall \phi (\neg \psi) \subseteq \Psi$

From $\neg \psi \in \Psi$, or the "easy direction": $M^c, s_\pi \models \neg \psi$.

Since $s_\pi, s_\psi \in R^c$, $M^c, s_\pi \models \neg \psi$, which contradicts the assumption.

Sahlqvist forms

Sahlqvist determined a general class of formulas C, such that any schema $\psi$ whose instantiations are in C (we say "it is in Sahlqvist form")

1. corresponds to a first-order definable class of frames,
2. whose logic can be axiomatized by adding $\psi$ to the axiomatization of the basic language.

Example: add $\diamond \psi \rightarrow \psi$ as an axiom for a sound and complete deductive system (K4) with respect to transitive frames.

Simple Sahlqvist forms (definition)

- **General Sahlqvist forms** have quite complicated definitions.
- In the course, we only need simple Sahlqvist forms.
- A positive formula is a modal formula where every atomic proposition is within the scope of an even number of negations
- Simple Sahlqvist forms: $\diamond \phi_1 \ldots \diamond \phi_k \ldots \square p \rightarrow B$
  where B is a positive formula.
The (deductive) systems $T, S_4, S_5, KD45$

- $T = K + \text{axiom } \Box \phi \rightarrow \phi$
- $S_4 = T + \text{axiom } (\Box \phi \rightarrow \Box \Box \phi)$
- $S_5 = S_4 + \text{axiom } \neg \Box \phi \rightarrow \neg \Box \Box \phi$
- $KD45 = K + \text{axioms } \{ \neg \bot, \phi \rightarrow \phi, \neg \Box \phi \rightarrow \neg \Box \Box \phi \}$

So, the basic system $K$ is extended with axioms corresponding to the first-order properties of the frames.

Decidability

- The satisfiability problem for all combinations of $D,T,B,4,5$ are decidable.
- The product $S_5 \times S_5 \times S_5$ is undecidable.

Complexities

- Recall:
  \[
  \text{NP} \subseteq \text{PSPACE} \subseteq \text{EXPTIME} \subseteq \text{NEXPTIME}
  \]
- $\text{NP}$-complete: satisfiability for $S_5, KD45, K45$
- $\text{PSPACE}$-complete: satisfiability for $K, KD, KT, K4, KD4, S4, S5_n, \text{LTL}$
- $\text{EXPTIME}$-complete: PDL, Common knowledge, global consequence in $K$
- $\text{NEXPTIME}$-complete: satisfiability for $S_5 \times S_5$

Weak modal logic

The well-known property $K$:

\[ \vdash \Box \psi \land (\psi \rightarrow \Box \psi) \rightarrow \Box \psi \]

This property is true in all (K)ripke models. If we want to get rid of it, we have to leave Kripke semantics. Some semantics weaker than Kripke semantics:

- neighborhood semantics,
- queer-world semantics,
- selection function semantics, etc.

Combining modal logics: Products

Given two frames $(U_0, R_0)$ and $(U_1, R_1)$, their product frame is $(U_0 \times U_1, R_{H}, R_{V})$, with $U_0 \times U_1$ the Cartesian product of the worlds from the originating models, and $R_H$ and $R_V$ accessibility relations defined by

1. $(w,x)R_H(y,z)$ if and only if $wR_0y$ and $x=z$, and
2. $(w,x)R_V(y,z)$ if and only if $xR_0z$ and $w=y$

Commutativity: $HV \phi \leftrightarrow VH \phi$

Confluence: $\neg H \neg V \phi \rightarrow V \neg H \phi$

Neighborhood models
Neighborhood models

A generally applicable semantics for weak modal logics.

\[
M = (W, V, N),
\]

- where \( W \) is again a non-empty set of states and \( V \) is a truth assignment function per state.
- \( N \) is a function \( W \to \mathcal{P}(\mathcal{P}(W) \setminus \{\emptyset\}) \); for every \( w, N(w) \) is a non-empty collection of non-empty subsets (neighborhoods) of \( W \).
- \( M, w \models \phi \iff \exists T \in N(w): t \in T \iff M, t \models \phi \)
- This semantics is quite weak...
- Extra conditions on neighborhoods make the semantics generally applicable. Closure under reachability of intersecting neighborhoods gives back \( \Box \phi \land \Box \psi \rightarrow \Box (\phi \land \psi) \), etc.

‘Applications’…
…of the modal theory to temporal, action, epistemic and deontic reasoning

Temporal logic

- \( X\phi = \) next time in future \( \phi \) (interpreted as a \( \Diamond \phi \))
- \( F\phi = \) some time in future \( \phi \) (interpreted as a \( \Diamond \phi \))
- \( G\phi = \) globally in the future \( \phi \) (interpreted as a \( \Box \phi \))

Possible axioms for ‘constructing’ a modal logic of time:

- \( FF\phi \rightarrow F\phi \) (transitivity)
- \( G\phi \rightarrow X\phi \)
- \( G\phi \equiv \phi \land XG\phi \)
- \( \phi \land G(\phi \rightarrow X\phi) \rightarrow G\phi \) (not Sahlqvist)

Often: two temporal dimensions: duration and choice

Dynamic logic

- In dynamic logics we have a separate accessibility relation for each action \( \Rightarrow \) multi-modal logic
- Simple version: accessibility is parameterized with respect to a set of atomic actions: \( [a]\phi \) means ‘performing action \( a \) guarantees that \( \phi \) is made true’
- More complex: in dynamic logic, actions have a structure:
  \( \alpha, \beta, ... ::= a \mid \phi \land \mid \alpha \lor \beta \lor \alpha \star \beta \)
  The action connectives have their usual relational interpretation.
  much of the previous theory is not applicable anymore...

Epistemic logic

- \( B\phi = \) in all worlds I deem possible according to what I believe, it holds that \( \phi \) (interpreted as a \( \Box \phi \))

Possible properties for ‘constructing’ a modal logic of belief:

- \( \neg(B\phi \land B\neg\phi) \) (belief consistency)
- \( B\phi \rightarrow BB\phi \) (positive introspection)
- \( \neg B\phi \rightarrow B\neg B\phi \) (negative introspection)

- KD45 most common for belief
- S5 for knowledge
- Groups: common knowledge
- In what sense can agents know an action?

The ‘epistemic logic’ of Donald Rumsfeld

Rumsfeld knows tautologies:

"We do know of certain knowledge that he [Osama Bin Laden] is either in Afghanistan, or in some other country, or dead."

Rumsfeld believes himself (abductively):

"I believe what I said yesterday. I don’t know what I said, but I know what I think, and, well, I assume it’s what I said."
Deontic Logic

- Obligation: $O \phi$ = in all worlds that are deontically optimal, it holds that $\phi$ ($O \phi$ interpreted as a $\Box \phi$)
- A key property of deontic logic is that $\neg \phi \land O \phi$ is consistent
- Permission: $P \phi \equiv_{def} \neg O \neg \phi$
- Prohibition: $F \phi \equiv_{def} \neg P \phi$
- Chisholm: $\phi \rightarrow O \psi$ nor $O(\phi \rightarrow \psi)$ express conditional obligation $\Rightarrow$ dyadic deontic logic

STIT Logic

- STIT is an acronym for 'Seeing To It That'
- Temporal logic of agency as opposed to a-temporal logics for 'Bringing It About That'.
- Central idea of all such logics: acting is ensuring the actual world is among a set of possible worlds ensured by performing the action.
- $[A] \phi$ means group A ensures that $\phi$
- In STIT this is applied to histories: acting is ensuring a certain set of histories.

What we saw today

- Modal logic has a well-developed mathematical basis.
- Modal logic is applicable to a wide range of reasoning domains.
- Modal logic has an appealing semantics, has nice properties, and is widely used.
- The basic modal system is well-suited to base more expressive logic theories on (agency, concurrency, communication, ...)

Outlook

- Tuesday: What agents can do
  - Logic of ability: Coalition Logic
- Wednesday: What agents do, what agents know they (can) do
  - Logic of agency: STIT
- Thursday: What agents want
  - logic of intention: Cohen and Levesque
- Friday: What agents can plan
  - ATL, Strategic STIT

Thanks for listening
See you tomorrow