An Introduction to Multiagent Systems

LECTURE 12: LOGICS FOR MULTIAgENT SYSTEMS
1 Overview

The aim is to give an overview of the ways that theorists conceptualise agents, and to summarise some of the key developments in agent theory.

- Study in agent theory
- Introduce the Cohen-Levesque theory of intention as a case
- Discuss Moore’s theory of ability
- Discuss modal logic as a tool for reasoning about attitudes
- Introduce some problems associated with formalising attitudes
- Characterise agents
- Discuss the various different attitudes that may be used to conceptualise agents

Begin by answering the question: Why theory?

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Formal methods have (arguably) had little impact on general software development, but progress cannot be made in language architectures, languages, and tools that we use — literally, a meaning. Without such a semantics, it is never clear exactly what is happening, or why it works.

End users (e.g., program managers) need never read or understand these semantics, but progress cannot be made in language architectures, languages, and tools that we use — literally, a meaning. Without such a semantics, it is never clear exactly what is happening, or why it works.

The answer is that we need to be able to give a semantics to the agent based systems? Why should they be relevant in practice of software development? Why should they be relevant in practice of software development? Why should they be relevant in practice of software development?

2 Why Theory?
In agent-based systems, we have a bag of concepts and tools, which are intuitively easy to understand (by means of metaphor and analogy), and have obvious potential. But we need theory to reach any kind of profound understanding of these tools.
Agents = Intentional Systems

Where do theorists start from?

The notion of an agent as an intentional system.

3 Agents = Intentional Systems

So agent theorists start with the (strong) view of agents as intentional systems: one whose simplest consistent description requires the intentional stance.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
We want to be able to design and build computer systems in terms of mentalistic notions. Before we can do this, we need to identify a tractable subset of these attitudes, and a model of how they interact to generate system behaviour.

So first, which attitudes?
Two categories:

- Information attitudes
  - Belief
  - Knowledge

- Pro-attitudes
  - Desire
  - Commitment
  - Obligation
  - Intention
  - Choice

Two categories:
Formalising Attitudes

Consider...

Janine believes Cronos is father of Zeus.

Naive translation into first-order logic:

\( \text{Bel}(\text{Janine, Father}(\text{Zeus, Cronos})) \)

- Intensional notions are referentially opaque.
- Consider \( \text{Zeus} = \text{Jupiter} \).
- Need to be able to apply 'Bel' to formulae of first-order logic, not a term.
- Allows us to substitute terms with the same denotation.
- No need to be able to apply 'Bel' to formulae of first-order logic, not a term.

So how do we formalise attitudes?

5 Formalising Attitudes
model logics, with possible worlds semantics.

• We will focus on modal languages, and in particular, normal

  that denote formulae of some other object-language.

  – use a meta-language: a first-order language containing terms
    which are applied to formulae;

  – use a modal language, which contains modal operators,

Two fundamental approaches to the syntactic problem:

•

Thus any formalism can be characterized in terms of two

  • a **semantic** one (no substitution of equivalents);
  • a **syntactic** one (intentional notions refer to sentences);

So, there are two sorts of problems to be addressed in developing

An Introduction to Multiagent Systems
We introduce a (propositional) modal logic for knowledge/belief.
Lecture 12: An Introduction to Multiagent Systems

**Syntax:**

- Modal connective: $K$
- Classical connectives: $\wedge, \vee, \neg$
- Primitive propositions: $\{ \ldots, a, b, c, d \} = \Phi$

**Vocabulary:**

- "knows that" operator $K$ is allowed.

**Example Formulae:**

So nesting of $K$ is allowed:

$$
\langle \text{ffm} \rangle K \\
\langle \text{ffm} \rangle \wedge \langle \text{ffm} \rangle \\
\langle \text{ffm} \rangle \neg \\
\Phi = \text{any member of } \langle \text{ffm} \rangle
$$

An Introduction to Multiagent Systems
\[ \frac{K(p \land q)}{K(p \land Kq)} \]
Semantics are trickier. The idea is that an agent's beliefs can be characterized as a set of possible worlds, in the following way.

An agent in a game such as poker, who possesses the ace of spades, could deduce what cards were held by her opponents. For example, any configuration in which she did not possess the ace of spades could be rejected. (For example, any configuration in which she did not possess the ace of spades could be rejected.)

First calculate all the various ways that the cards in the pack could possibly have been distributed among the various players. The systematically eliminate all those configurations which are not possible, given what she knows. The systematically eliminate all those configurations which are not possible, given what she knows.

An agent playing a card game such as poker, who possesses the ace of spades, could deduce what cards were held by her opponents. How could she deduce what cards were held by her opponents?
Each configuration remaining after this is a world; a state of affairs considered possible, given what she knows.

Something true in all our agents' possibilities is believed by the agent.

For example, in all our agent's epistemic alternatives, she has the ace of spades. Two advantages:

- remains neutral on the cognitive structure of agents;
- the associated mathematical theory is very nice!

http://www.csc.liv.ac.uk/~mjw/pubs/imas/13
To formalise all this, let $W$ be a set of worlds, and let $R \subseteq W \times W$ be a binary relation on $W$, characterising what worlds the agent considers possible.

Semantic of formulae are given relative to worlds: in particular:

- $K \phi$ is true in world $w$ iff $\phi$ is true in all worlds $w'$ such that $(w', w) \in R$.

For example, if $(w, w') \in R$, then if the agent was actually in world $w'$, the agent might believe it was in world $w$.

To formalise all this, let $W$ be a set of worlds, and let $R \subseteq W \times W$ be a binary relation on $W$, characterising what worlds the agent considers possible.
This is not a desirable property.
This is *logical omniscience*.

Thus agent’s knowledge is closed under logical consequence:

- If $\phi$ is valid, then $\phi K$ is valid.

  $$ (\phi Y \iff \phi K) \iff (\phi \iff \phi) K $$

- The following axiom schema is valid:

  Two basic properties of this definition:
The most interesting properties of this logic turn out to be those relating to the properties we can impose on accessibility relation \( R \).

By imposing various constraints, we end up getting out various axioms; there are lots of these, but the most important are:

\[
\begin{align*}
\phi \iff \phi K \K \phi K & \quad 5 \\
\phi KK \iff \phi K & \quad 4 \\
\phi \K \phi K & \iff \phi K & \quad D \\
\phi \iff \phi K & \quad T
\end{align*}
\]

The most interesting properties of this logic turn out to be those

An Introduction to Multiagent Systems
Axiom T is the knowledge axiom: it says that what is known is true.

Axiom D is the consistency axiom: if you know φ, you can’t also know ¬φ.

Axiom 4 is positive introspection: if you know φ, you know you know φ.

Axiom 5 is negative introspection: you are aware of what you don’t know.

We can (to a certain extent) pick and choose which axioms we want to represent our agents.

All of these (KTD45) constitute the logical system S5.

S5 without the T is weak-S5, or KD45.

Often chosen as a logic of idealised knowledge.

(http://www.csc.liv.ac.uk/~mjw/pubs/tmas/)
Lecture 12
An Introduction to Multiagent Systems

Most-studied aspect of practical reasoning agents:

Knowledge & Action

Moore's 1977 analysis is best-known in this area.
Formal tools:
- modal logic
- Kripke semantics
- dynamic logic

Moore showed how Kripke semantics could be axiomatized in a

First-order meta-language:
- model formulae then translated to meta-language using

representation for action;
- a modal logic with Kripke semantics + dynamic logic-style

interaction between knowledge and action.

Model theorem proving reduces to meta-language theorem

Axiomatization;

- modal theorem proving reduces to meta-language theorem

Theorem proving:
- modal tools:

Interaction between knowledge and action.

Most-studied aspect of practical reasoning agents:

Knowledge & Action

http://www.csc.liv.ac.uk/~mjw/pubs/tmas/
Moore considered 2 aspects of interaction between knowledge and action:

1. As a result of performing an action, an agent can gain knowledge. Agents can perform "test" actions, in order to find things out.
2. In order to perform some actions, an agent needs knowledge: these are knowledge pre-conditions. For example, in order to open a safe, it is necessary to know the combination.

Culminated in defn of ability: what it means to be able to do bring something about.
Axiomatising standard logical connectives:

- True is a meta-language predicate:

Here, \( \text{True} \) is a meta-language predicate:

\[
\begin{align*}
((\phi, w)_{\text{True}} & \iff (\phi, w)_{\text{True}}) \iff (\phi \iff \phi, w)_{\text{True}} \cdot \text{M} \\
(\phi, w)_{\text{True}} & \iff (\phi, w)_{\text{True}} \iff (\phi \iff \phi, w)_{\text{True}} \cdot \text{M} \\
(\phi, w)_{\text{True}} & \iff (\phi, w)_{\text{True}} \iff (\phi \land \phi, w)_{\text{True}} \cdot \text{M} \\
(\phi, w)_{\text{True}} & \iff (\phi, w)_{\text{True}} \iff (\phi \lor \phi, w)_{\text{True}} \cdot \text{M} \\
(\phi, w)_{\text{True}} & \iff (\phi, w)_{\text{True}} \iff (\phi, w)_{\text{F}} \cdot \text{M} \\
\end{align*}
\]

Frege quotes, ‘\( \lceil \cdot \rceil \)’, used to quote modal language formula.

- 1st argument is a term denoting a world;
- 2nd argument is a term denoting modal language formula.

An Introduction to Multiagent Systems

http://www.csc.liv.ac.uk/~mjw/pubs/imas/20
Axiomatizing the knowledge connective: basic possible worlds

Here, $K$ is a meta-language predicate used to represent the knowledge accessibility relation.

\[
([\phi], m) \wedge \text{true} \iff (m, m) \mathcal{K} \cdot m = (([\text{know}(\phi)], m) \wedge \text{true})
\]

Semantics:
These axioms ensure that \( K \) is an equivalence relation.

\[
\left( \left( \mathcal{M}, \mathcal{M} \right) K \right) \iff \left( \mathcal{M}, \mathcal{M} \right) K \lor \left( \mathcal{M}, \mathcal{M} \right) K \land \neg \left( \mathcal{M}, \mathcal{M} \right) K
\]

Euclidean:

\[
\left( \mathcal{M}, \mathcal{M} \right) K \iff \left( \mathcal{M}, \mathcal{M} \right) K \lor \left( \mathcal{M}, \mathcal{M} \right) K \land \neg \left( \mathcal{M}, \mathcal{M} \right) K
\]

Transitive:

\[
\left( \mathcal{M}, \mathcal{M} \right) K
\]

Reflexive:

\[
\left( \mathcal{M}, \mathcal{M} \right) K
\]

Other axioms added to represent properties of knowledge.
Now we need some apparatus for representing actions.

- Second says that a necessary consequence of performing an action is possible:

\[ ([\phi] \text{True}, w) \iff ([w, a, w]) R \cdot M \land \forall ([w, a, w]) R \cdot M \]

- First conjunct says the action is possible:

\[ ([\phi] \text{True}, w) \iff ([w, a, w]) R \cdot M \land \forall ([w, a, w]) R \cdot M \]

Then introduce a modal operator (Res a) to mean that after action a performed, action a will be true.

- Add a meta-language predicate R(a, w) to mean that w is a world that could result from performing action a in world w.

An Introduction to Multiagent Systems
Now we can define ability, via modal can operator:

\[
[\langle \phi \; \text{can} \rangle]_w^{\text{true}} \iff [\langle \phi \; \text{can} \rangle]_w^{\text{true}} \wedge \text{true}
\]

Terminology: $\nu$ is quantified de re.

Has a "definite description" of it.

Implies agent knows the identity of the action.

Note the way $\nu$ is quantified w.r.t. the Know modality.

$\phi$ is such that the result of performing $\nu$ is $\phi$.

So agent can achieve $\phi$ if there exists some action $\nu$, such that agent knows that the result of performing $\nu$ is $\phi$.

agent knows the identity of the action. 

$a$ is quantified dere.

Has a "definite description" of it.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
A circular definition?

No, interpret as a fixed point.

We can weaken the definition to capture the case where an agent performs an action to find out how to achieve a goal.
Lecture 12 An Introduction to Multiagent Systems

Critique of Moore's formalism:

1. Translating modal logic (int. modal & dynamic logic) to bear on rational agency to first serious attempt to use tools of mathematical logic (incl. modal & dynamic logic) to bear on rational agency.

- But probably first serious attempt to use tools of mathematical logic (incl. modal & dynamic logic) to bear on rational agency.

Definition of ability is somewhat vacuous.

Logical omniscience.

3. Moore's formalism based on possible worlds: fails prey to logical omniscience.

Original structure (and hence sense) is lost.

Complicated and unintuitive.

2. Formulae resulting from the translation process are more efficient.

"Hard-wired" modal theorem provers will be more efficient.

Theorem proving in first-order language is inefficient.

1. Translating modal language into a first-order one and then theorem proving in first-order language is inefficient.

Critique of Moore's formalism:

http://www.csc.liv.ac.uk/~mjw/pubs/tmas/
We have one aspect of an agent, but knowledge/belief alone does not completely characterize an agent. We need a set of connectives, for talking about an agent's pro-attitudes as well.

Here, we review one attempt to produce a coherent account of how the components of an agent's cognitive state hold together: the theory of intention developed by Cohen & Levesque. The theory of intention is the study of how the components of an agent's cognitive state hold together. It is my intention to prepare my slides.

We need a set of connectives, for talking about an agent's pro-attitudes as well. We have one aspect of an agent, but knowledge/belief alone does not completely characterize an agent.
8.1 What is intention?

Two sorts:

- Future directed
  - To serve to coordinate future activity.
  - To attitude to a proposition
  - Future directed
  - To function causally in producing behavior.
  - To attitude to an action
  - Present directed

We are here concerned with future directed intentions.
Following Bratman (1987) Cohen-Levesque identify seven properties that must be satisfied by intention:

1. Intentions pose problems for agents, who need to determine ways of achieving them.

2. Intentions provide a filter for adopting other intentions, which resources to devoting how to bring about.

3. Agents track the success of their intentions, and are inclined to try again if their attempts fail.

4. Intentions pose problems for agents, who need to determine ways of achieving them.

5. Intentions must not conflict.

6. Intentions provide a filter for adopting other intentions, which are mutually exclusive.

7. If an agent's first attempt to achieve φ fails, then all other things try again if their attempts fail.

If I have an intention to φ, you would expect me to adopt an intention such that φ and φ are mutually exclusive.

If an agent has an alternative plan to achieve φ, it will try an alternative plan to achieve φ.

Following Cohen-Levesque identity seven properties that must be satisfied by intention:

An Introduction to Multiagent Systems
Agents believe their intentions are possible. (CTL notation: $\phi \square \phi$)

Agents do not believe they will not bring about their intentions.

It would not be rational of me to adopt an intention to if I believed was not possible. (CTL* notation: $\neg \neg \phi$)

Agents believe their intentions are possible. (CTL* notation: $\neg \neg \phi$)

Agents do not believe they will not bring about their intentions.

That is, they believe there is at least some way that the intentions could be brought about. (CTL notation: $\exists \phi$)

In addition...
An Introduction to Multiagent Systems

Under certain circumstances, agents believe they will bring about their intentions.

While this does not imply that I intend to suffer pain if I go to the dentist — but this does not make sense if I believe is inevitable (CTL*: $\phi \Rightarrow \text{A}$) that I would adopt it as an intention.

Moreover, it does not make sense that if I believe is inevitable (CTL*: $\phi \Rightarrow \text{A}$) that I would bring my intentions about; intentions can fail. Nevertheless, it does bring my intentions about; intentions can fail. Moreover, it does not normally be rational of me to believe that I would bring about their intentions. If I believe is inevitable (CTL*: $\phi \Rightarrow \text{A}$) that I would adopt it as an intention.

Moreover, it does not make sense that if I believe is inevitable (CTL*: $\phi \Rightarrow \text{A}$) that I would bring my intentions about; intentions can fail. Nevertheless, it does bring about their intentions. If I believe is inevitable (CTL*: $\phi \Rightarrow \text{A}$) that I would adopt it as an intention.
Cohen-Levesque use a multi-modal logic with the following major

- belief accessibility relation — $B$

- Each agent allocated:
- Each world is infinitely long linear sequence of states.
- Semantics are possible worlds.

$$\begin{align*}
\text{Done} & \quad (\alpha) \\
\text{Happens} & \quad (\alpha)
\end{align*}$$

$$\begin{align*}
\phi & \quad \text{has goal of} \\
\phi & \quad \text{believes}
\end{align*}$$

for every agent/time pair, gives a set of belief accessible worlds;

euclidean, serial, transitive — gives belief logic K45.

Semantics are possible worlds.

An introduction to Multiagent Systems
- goal accessibility relation \( G \) for every agent/time pair, gives a set of goal accessible worlds.

Serial — gives goal logic KD.
A constraint: $G \not\subseteq B$.

- A realism property — agents accept the inevitable:
  $$ (\phi \not\models (G\models \phi) ) \models (\phi \not\models (\Diamond (G\models \phi))) $$

- Another constraint:

- A constraint: $G \subseteq C$.

$C$ is a constraint assumption capturing the following properties:
- agents do not indefinitely defer working on goals;
- agents do not persist with goals forever.
Add some operators for describing the structure of event sequences followed by 'test action', 'always' and 'sometimes' can be defined as abbreviations, along with a "strict" sometime operator, 'later':

\[
\begin{align*}
d\lozenge \lor d\Box & \equiv \text{ (later) } d \\
\forall x \lozenge \Box x & \equiv x \Box \\
(\exists \text{ happens } x; \forall x) \Xi & \equiv \forall x \lozenge
\end{align*}
\]

Also add some operators of temporal logic, "always" (\(\Box\)) and "sometimes" (\(\lozenge\)) test action, \(\forall a\) followed by \(\forall a\); \(a\); \(a\), and in some operators for describing the structure of event sequences of multiagent systems.
Finally, a temporal precedence operator, \( \text{before} \).
– the agent believes the goal will never be satisfied;

– the agent believes the goal has been satisfied;
will not be paid, the intention evaporates...

Next, intention:

A single agent has an intention to do a if it has a persistent goal and it has an intention to do a, and then done a.

\[
\\forall x \left( \text{Goal} - \text{Goal} \Rightarrow \text{Intend} \left( x \wedge \text{Goal} \right) \right)
\]

Main point: avoids ever committing.

C&L discuss how this definition satisfies desiderata for intention.

Adaptation of definition allows for relativised intentions. Example:

So, an agent has an intention to do a if it has a persistent goal.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/38
Critique of C&L theory of intention (Singh, 1992):

- does not capture and adequate notion of “competence”;
- does not adequately represent intentions to do composite actions;
- requires that agents know what they are about to do — fully elaborated intentions;
- disallows multiple intentions.
Lecture 12: An Introduction to Multiagent Systems

9 Semantics for Speech Acts

C&L used their theory of intention to develop a theory of several speech acts.

First, define alternating belief.

We will look at request.

These actions.

C&L use their dynamic logic-style formalism for representing actions.

Key observation: illocutionary acts are complex event types (ct).

C&L used their theory of intention to develop a theory of several speech acts.
And the related concept of mutual belief:

\[ (d \land x \land u \land \text{Bel})(A) \land \text{Bel}(d) \land x \land u \land \text{Bel}(M - \text{Bel}(M)) \]
An attempt is defined as a complex action expression by doing $d$ - represents ultimate goal that agent is aiming for by doing $e$,' order of operations determine meaning:

Here:

In English: "An attempt is a complex action that agents perform when they do something ($e$) desiring to bring about some effect ($q$) but with intent to produce at least some result ($p$)."

\[
\text{\texttt{Attempt}} \equiv \{ b \mid \forall \{d \mid \forall \{e \mid \forall \{x \mid \text{\texttt{Goal}}(x) \land \text{\texttt{Happens}}(x, e) \land \text{\texttt{Intends}}(x, e) \land \text{\texttt{At}e\texttt{mp}t}(e) \}} \}
\]

(\text{\texttt{model operator}})

Hence the use of curly brackets, to distinguish from predicate or atom. 

An attempt is defined as a complex action expression.
Proposition $q$ represents what it takes to at least make an “honest effort” to achieve $d$. Proposition $b$ represents what it takes to at least make an
In English:

\[
\begin{align*}
&\left( \Diamond \text{Goal}_x \right) \iff \\
&\left( \square \left( \text{Goal}_x \land \Box \text{Done}_x \right) \right) \land \\
&\forall e \left( \left( \text{Bel}_x \left( \Diamond \text{Goal}_x \right) \land \exists y \left( \text{Helpful}_y \right) \right) \implies \text{Helpful}_x \right)
\end{align*}
\]

Definition of helpfulness needed: •
Definition of requests:

\[
\begin{align*}
\text{Request} &: a \land (\text{Goal}<\text{spkr}> \lor \text{Done}<\text{spkr}> \lor \text{Intend}<\text{addr}> \lor \text{Done}<\text{addr}> \\
\text{Attempt} &: e \land (\text{Goal}<\text{spkr}> \lor \text{Done}<\text{spkr}> \lor \text{Intend}<\text{addr}> \lor \text{Done}<\text{addr}>)
\end{align*}
\]

In English:

\[
\begin{align*}
\text{Request} &= a \land (\text{Helpful}<\text{spkr}> \lor \text{Goal}<\text{spkr}> \lor \text{Done}<\text{spkr}> \lor \text{Intend}<\text{addr}> \lor \text{Done}<\text{addr}> \\
\text{Attempt} &= e \land (\text{Goal}<\text{spkr}> \lor \text{Done}<\text{spkr}> \lor \text{Intend}<\text{addr}> \lor \text{Done}<\text{addr}>)
\end{align*}
\]

\[
\begin{align*}
\text{Where} \\
\phi &= \{ M - \text{Bel}\text{<}\text{addr}<\text{spkr}> (\text{Goal}<\text{spkr}>) \lor \text{Attempt}<\text{spkr}> e \} \\
\Rightarrow &\{ \text{Request}<\text{spkr}> e \lor a \}
\end{align*}
\]
A request is an attempt on the part of the speaker, by doing e, to bring about a state where ideally 1) the addressee intends, relative to the speaker still having that goal, and addressee still being helpfully inclined to the speaker, and 2) addressee actually eventually brings about a state where ideally, addressee by doing e, to bring about the ideal situation.

It is mutually believed that if wants the ideal situation does e, or at least brings about a state where ideally addressee believes the ideal situation.
By this definition, there is no primitive request act.

[A] speaker is viewed as having performed a request if he executes any sequence of actions that produces the needed effects.
10 A Theory of Cooperation

- We now move on to a theory of cooperation (or more precisely, cooperative problem solving).
- This theory draws on work such as C&L’s model of intention, and their semantics for speech acts.
- It uses connectives such as ‘intend’ as the building blocks.
- The theory intends to explain how an agent can start with an desire, and be moved to get other agents involved with achieving this desire.

Wenowmoveontoatheoryofcooperation(ormoreprecisely,cooperativeproblemsolving).

ThistheorydrawsonworksuchasC&L’smodelofintention,andeithersemanticsforspeechacts.

Itusesconnectives suchas‘intend’asthebuildingblocks.

Thetheoryintendstoexplainhowanagentcanstartwithan
desire,andebeamovetogetotheragentsinvolvedwithachieving
thisdesire.
We formalise our theory by expressing it in a quantified multi-modal logic.

- beliefs;
- goals;
- actions (transitions in branching time structure) associated with agents;
- groups (sets of agents) as terms in the language — set theoretic mechanism for reasoning about groups;
- path quantifiers (branching time);
- dynamic logic style action constructors;

We formalise our theory by expressing it in a quantified multi-modal logic.

Another ( Formal Framework

Formal semantics in the paper.
1. Recognition.

CPS begins when some agent recognises the potential for cooperative action at stage (1) solicit assistance.

The agent that recognised the potential for cooperative action at stage (1) solicits assistance. May happen because an agent has a goal that it is unable to achieve in isolation, or because the agent prefers assistance.

If team formation successful, then it will end with a group having a joint commitment to collective action.

2. Team formation.

CPS begins when some agent recognises the potential for cooperative action.

An Introduction to Multiagent Systems
3. Plan formation. The agents attempt to negotiate a joint plan that they believe will achieve the desired goal.

4. Team action. The newly agreed plan of joint action is executed by the agents, which maintain a close-knit relationship throughout.
Recognition begins when some agent has a goal, and recognizes the potential for cooperative action with respect to that goal.

CPS typically begins when some agent has a goal, and recognizes that goal.

Recall that recognition may occur for several reasons:

- The goal is unachievable in isolation, due to lack of resources, but believes that cooperative action can achieve it.
- An agent may have the resources to achieve the goal but does not want to use them.
- An agent is unable to achieve its goal in isolation, due to a lack of resources, but believes that cooperative action can achieve it.
- The agent is unable to achieve the goal, but recognizes the potential for cooperative action with respect to that goal.
Lecture 12: An Introduction to Multiagent Systems

- Doesn’t mean it doesn’t happen next.
  \( \phi'[a] \)
- Achieves a is dynamic logic
  \(\text{Achieves } a \)
- Can is a generalization of Moore’s
  \( \exists ! \text{ Agent } \cdot \text{Bel} \)
- Can is essentially Moore’s
  \( \land (\phi \land \text{Can}) \)

Note:

\[
\begin{aligned}
\left[ \left( \left( \neg \text{Goal } \right) \right) \right. \\
\iff \left( \phi \land \text{Achieves } a \right) \\
\land \left( \text{Bel } \lor \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\land \left( \phi \land \text{Can} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\lor \left( \phi \land \text{Agent } \cdot \text{Bel} \lor \text{Agent } \cdot \text{Bel} \right) \\
\end{aligned}
\]

Formally...
Having identified the potential for cooperative action with respect to one of its goals, a rational agent will solicit assistance from a group of agents that it believes can achieve the goal. If the agent is successful, then it will have brought about a mental state wherein the group has a joint commitment to collective action. Note that an agent cannot guarantee that it will be successful in forming a team; it can only attempt it.

12.2 Team Formation
- J - Commit is similar to J - P - Goal.
- Team is defined in later;

Note that:

\[
(\phi \& \text{Team} \& (J - \text{Can}) \& (J - \text{Bel})) \equiv (I \phi \& \text{PreTeam})
\]

Formally...
The main assumption concerning team formation can now be stated.

\[
\begin{align*}
&\left(\left( (\phi \land \text{Team}) \land (\phi \land \text{Goal}) \right) \lor (\phi \land \text{Bel} \land \text{Goal}) \right) \Rightarrow b \\
&\left( (\phi \land \text{Team}) \right) \Rightarrow d
\end{align*}
\]

where

\[
\begin{align*}
&\left( \{b, d \land \text{Attempt}\} \land \text{Happens} \right) \Rightarrow \\
&\forall E \exists \Diamond \forall \text{Bel} \land \text{Potential} - \text{for} - \text{Bel} \land \phi \land \text{Bel} \land \text{Goal} \Rightarrow \text{Goal}
\end{align*}
\]
If team formation is successful, then there will be a group of agents with a joint commitment to collective action. But collective action cannot begin until the group agrees on what they will actually do. Hence the next stage in the CPS process: plan formation, which involves negotiation.

Unfortunately, negotiation is extremely complex — we simply offer some observations about the weakest conditions under which negotiation can be said to have occurred.

\[12.3 \text{ Plan Formation}\]
Lecture 12: An Introduction to Multiagent Systems

Note that negotiation may fail: the collective may simply be unable to reach agreement.

If negotiation succeeds, we expect a team action stage to follow.

- If negotiation succeeds, we expect a team action stage to follow.
- Collective closer to the goal.
- Proposed a course of action that it believed would take the collective closer to the goal.
- Say that negotiation occurred at all is that at least one agent proposed a course of action that it believed would take the collective closer to the goal.
- In this case, the minimum condition required for us to be able to say that negotiation occurred at all is that at least one agent proposed a course of action that it believed would take the collective closer to the goal.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
We might also assume that agents will attempt to bring about their preferences. For example, if an agent has an objection to some plan, then it will attempt to prevent this plan being carried out.

The main assumption is then:

We might also assume that agents will attempt to bring about their preferences.
Team action simply involves the team jointly intending to achieve the goal. The formalisation of \( \text{Team} \) is simple.

\[
(\phi \land \text{Intend} \phi) \\
\land \exists (\phi \land \text{Achieves} \phi) \rightarrow (\phi \land \text{Team})
\]