An Introduction to Multiagent Systems

CHAPTER 17: LOGICAL FOUNDATIONS

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Overview

• Introduce modal logic as a tool for reasoning about attitudes.

• Discuss the various different attitudes that may be used to characterise agents.

• Begin by answering the question: Why theory?

• Introduce some problems associated with formalising the theorists conceptualise agents, and to summarise some of the key developments in agent theory.

• The aim is to give an overview of the ways that theorists conceptualise agents.
Discuss Moore’s theory of ability.

Introduce the Cohen-Levesque theory of intention as a case study in agent theory.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Chapter 17

An Introduction to Multiagent Systems

2 Why Theory?

2 Why Theory?

Formal methods have (arguably) had little impact on the general practice of software development. Why should they be relevant in agent-based systems?

What is happening, or why it works.

Without such a semantics, it is never clear exactly that we use — literally, a meaning.

The answer is that we need to be able to give a semantics to the architectures, languages, and tools.

• Why Theory?

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
End users (e.g., programmers) need never read or understand the system semantics, but progress cannot be made in language development until those semantics exist. In agent-based systems, we have a bag of concepts, 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.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
consistent description requires the intentional stance: agents as intentional systems: one whose simplest consistent description requires the intentional stance.

So agent theorists start with the (strong) view of

The notion of an agent as an intentional system.

Where do theorists start from?

3 Agents = Intentional Systems

An Introduction to Multiagent Systems 2e
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.

4 Theories of Attitudes

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Some possibilities:

- Information attitudes
  - Belief
  - Knowledge

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

- So how do we formalise attitudes?

Janine believes Cronos is father of Zeus.

Naive translation into first-order logic:

\[ Bel(Janine, Father(Zeus, Cronos)) \]

- But...

- The second argument to the Bel predicate is a formula of first-order logic, not a term.

- So how do we formalise attitudes?
intentional notions are referentially opaque.

(Zeus = Jupiter)

denotation: consider substitution terms with the same
allows us to substitute terms with the same
So, there are two sorts of problems to be addressed in developing a logical formalism for intentional notions:

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

Thus any formalism can be characterized in terms of two attributes: its *language of formulation*, and *semantic model*:

Two fundamental approaches to the syntactic problem:
normal modal logics, with possible worlds semantics.
• We will focus on modal languages, and in particular,
other object-language containing terms that denote formulae of some
- use a meta-language: a first-order language
  operators, which are applied to formulae,
- use a modal language, which contains modal
6 Normal Modal Logic for Knowledge
Syntaxis

Vocabulary

Syntax:

modal connective

classical connectives

primitive propositions

\[ \Phi = \{ p, q, r, \ldots \} \]

\[ \wedge, \vee, \neg \]

\[ K \]

\[ \vdash, \models, \Rightarrow, \Leftrightarrow \]

\{ \ldots, a, b, c, d \} = \Phi

Syntaxis is classical propositional logic, plus an

operator K for 'knows that'.
Example formulae:

\[ (b \land q) \]

\[ (p \land K q) \]
The various players.

The pack could possibly have been distributed among the various ways that the cards in the pack could possibly have been distributed among the various players.

First calculate all the various ways that the cards in the pack could possibly have been distributed among the various players.

How could she deduce what cards were held by her opponents?

Consider an agent playing a card game such as poker, who possess the ace of spades.

Consider an agent playing a card game such as poker, who possess the ace of spades.

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 Introduction to Multiagent Systems 2e
Chapter 17
The systematically eliminate all those configurations which are not possible, given what she knows. (For example, any configuration in which she did not possess the ace of spades could be rejected.)
– the associated mathematical theory is very nice!
– remains neutral on the cognitive structure of agents;

Two advantages:

• she has the ace of spades.
For example, in all our agent's epistemic alternatives, something true in all our agent's possibilities is believed by the agent.

• knows.
Each configuration remaining after this is a world; a state of affairs considered possible, given what she

An Introduction to Multiagent Systems 2e
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. For example, if $(w, w') \in R$, then in the agent was actually in world $w$, then as far as it was concerned, it might be in world $w'$. The semantics of formulae are given relative to worlds: in particular:

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

For example, $(w, w') \in R$.

Chapter 17
An Introduction to Multi-agent Systems 26
http://www.csc.liv.ac.uk/~mjw/pubs/imas/
This is not a desirable property.

Thus agents' knowledge is closed under logical consequence: this is logical omniscience. This is not a desirable property!

- If $\phi$ is valid, then $\Box \phi$ is valid.
  
  $$ (\phi \iff \Box \phi) \iff (\phi \iff \phi) $$

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 lots of these, but the most important are:

$$\phi K \iff \exists K \phi$$

$$\phi KK \iff \phi K \iff \phi$$

$$\phi \iff \phi K$$

$$\bot$$

An Introduction to Multiagent Systems 2e
• **Axiom T** is the knowledge axiom: what you don’t know.

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

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

• **Axiom 3** is the consistency axiom: if you know φ, you can’t also know ¬φ.

• **Axiom D** is the consistency axiom: known is true.

• **Axiom T** is the knowledge axiom: it says that what is known is true.

**Interpreting the Axioms**
We can (to a certain extent) pick and choose which axioms we want to represent our agents. Often chosen as a logic of idealised knowledge. Often chosen as a logic of idealised belief. S5 without T is weak-S5, or KD45. All of these (KTD45) constitute the logical system S5. Systems of Knowledge & Belief.
Most-studied aspect of practical reasoning agents: Knowledge & Action

Chapter 17

An Introduction to Multiagent Systems

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
modal theorem proving reduces to meta-language.

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

1. As a result of performing an action, an agent can gain knowledge.
2. In order to perform some actions, an agent needs knowledge: these are knowledge pre-conditions.

Culminated in definition of ability: what it means to be able to do brings something about.

For example, in order to open a safe, it is necessary to know the combination.

Agents can perform "test" actions, in order to find things out.

Culminated in definition of ability: what it means to be able to do brings something about.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/25
Axiomatising standard logical connectives:

\[ \forall w. \text{True}(w, \lceil \neg \phi \rceil) \iff \neg \text{True}(w, \lceil \phi \rceil) \]

\[ \forall w. \text{True}(w, \lceil \phi \land \psi \rceil) \iff \text{True}(w, \lceil \phi \rceil) \land \text{True}(w, \lceil \psi \rceil) \]

\[ \forall w. \text{True}(w, \lceil \phi \lor \psi \rceil) \iff \text{True}(w, \lceil \phi \rceil) \lor \text{True}(w, \lceil \psi \rceil) \]

\[ \forall w. \text{True}(w, \lceil \phi \rightarrow \psi \rceil) \iff \text{True}(w, \lceil \phi \rceil) \rightarrow \text{True}(w, \lceil \psi \rceil) \]

\[ \forall w. \text{True}(w, \lceil \phi \leftrightarrow \psi \rceil) \iff \text{True}(w, \lceil \phi \rceil) \leftrightarrow \text{True}(w, \lceil \psi \rceil) \]

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

\[ (\lceil \phi \rceil, \text{True}) \iff (\lceil \phi \rceil, \text{True}) \iff (\lceil \phi \iff \phi \rceil, \text{True}) \]

\[ (\lceil \phi \rceil, \text{True}) \iff (\lceil \phi \rceil, \text{True}) \iff (\lceil \phi \land \phi \rceil, \text{True}) \]

\[ (\lceil \phi \rceil, \text{True}) \iff (\lceil \phi \lor \phi \rceil, \text{True}) \]

\[ (\lceil \phi \rceil, \text{True}) \iff (\lceil \neg \phi \rceil, \text{True}) \]

Axiomatising standard logical connectives:

An Introduction to Multiagent Systems
Fregula's quotes, \[ \langle \cdot \rangle \text{ used to quote modal language formula.} \]
Axiomatizing the knowledge connective: basic

Possible world semantics:

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

- Reflexive:
  $$\forall m \cdot K(m, m)$$

- Transitive:
  $$\forall m, m', m'' \cdot K(m, m') \land K(m', m'') \Rightarrow K(m, m'')$$

- Euclidean:
  $$\forall m, m', m'' \cdot K(m, m') \land K(m', m'') \Rightarrow K(m, m'')$$

Ensures that $K$ is an equivalence relation.

Other axioms added to represent properties of knowledge.

Axiomatizing the knowledge connective: basic
Now we need some apparatus for representing actions.

- First conjunct says the action is possible:

\[
\left[\phi\right]_w^{\text{true}} \iff \left(\frac{\text{true}}{w}\right)^v \cdot \frac{\text{true}}{w} \\lor \left(\frac{\text{true}}{w}\right)^v \cdot \frac{\text{true}}{w} \Rightarrow \left[\phi \land \text{a} \right]_w^{\text{true}}
\]

- Then introduce a modal operator \(\text{Res} a \phi\) to mean that after action \(a\) is performed, \(\phi\) will be true.

- Action \(a\) in world \(w\).

- Add a meta-language predicate \(R(a, w, w')\) to mean that \(w'\) is a world that could result from performing action \(a\) in world \(w\).
- second says that a necessary consequence of performing action is $\phi$. 

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Now we can define ability, via modal can operator.

\( \forall w \cdot \text{True} (w, \lfloor \text{Can} \phi \rfloor) \iff \exists a. \text{True} (w, \lfloor \text{Know} (\text{Res} a \phi) \rfloor) \)

So agent can achieve \( \phi \) if there exists some action \( a \) such that agent knows that the result of performing \( a \) is \( \phi \) and the agent knows the identity of the action. Note the way \( \forall \) is quantified w.r.t. the know modality. Has a “definite description” of it. Imply agent knows the identity of the action. Implies agent knows the identity of the action.

Terminology: \( a \) is quantified de re.

(\( a \) is quantified de re.)
Wecanweakenthedefinition,tocapturethecase
whereanagentperformsanactiontofindouthowto
achievegoal.

\[ \forall w \cdot \text{True}(w, \left[ \left[ \text{Can}\phi\right] \right](w)) \iff \exists a. \text{True}(w, \left[ \left[ \text{Know}(\text{Res}a\phi)\right] \right](w)) \lor \exists a. \text{True}(w, \left[ \left[ \text{Know}(\text{Res}a(\text{Can}\phi))\right] \right](w)) \]

No, interpret as a fixed point.

A circular definition?

Acirculardefinition?

AnIntroductiontoMultiagentSystems2e

Chapter17
Chapter 17: An Introduction to Multiagent Systems

Critique of Moore's formalism:

1. Translating modal language into a first-order one is inefficient.

2. Formulae resulting from the translation process are complicated and unintuitive.

3. Moore's formalism based on possible worlds fails. Original structure (and hence sense) is lost.

Definition of ability is somewhat vacuous.

...to logical omniscience.

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

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
But probably first serious attempt to use tools of mathematical logic (incl. modal & dynamic logic) to bear on rational agency.
• We have one aspect of an agent, but knowledge/belief alone does not completely characterise an agent.

• We need a set of connectives, for talking about an agent's pro-attitudes as well.

• We need to achieve a rational balance between its attitudes:
  - shouldn't be over-committed;
  - shouldn't be under-committed.

Agent's intention

8 Intention
It is my intention to prepare my slides.

Here we mean intention as in...

Developed by Cohen & Levesque: cognitive state hold together: the theory of intention account of how the components of an agent's

Here, we review one attempt to produce a coherent

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Chapter 17  An Introduction to Multiagent Systems

8.1 What is intention?

We are here concerned with future directed intentions.

- present directed
  - attitude to an action
  - function causally in producing behaviour.
- future directed
  - attitude to a proposition
  - serve to coordinate future activity.

Two sorts:

- future directed
- present directed
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.
   
   *If I have an intention to $\phi$, you would expect me to devote resources to deciding how to bring about $\phi$."

2. Intentions provide a ‘filter’ for adopting other intentions, which must not conflict.
   
   *If I have an intention to $\phi$, you would expect me to adopt an intention $\psi$ such that $\phi$ and $\psi$ are mutually exclusive."
An Introduction to Multiagent Systems

Chapter 17

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

If an agent's first attempt to achieve $\phi$ fails, then all other things being equal, it will try an alternative plan to achieve $\phi$. If $\phi$ to achieve $\phi$.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
In addition...

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

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

If I believed $\phi$ was not possible, it would not be rational of me to adopt an intention to bring about its intentions. (CTL* notation: $A
\square \neg \phi$).

In addition...

• Agents believe their intentions are possible.
• Under certain circumstances, agents believe they will bring about their intentions.

Moreover, it does not make sense that I believe $\phi$ is inevitable (CTL*: $A^{\Diamond} \phi$) that I would adopt it as an intention. It would bring my intentions about; intentions can fail. It would not normally be rational of me to believe that I would adopt my intentions.
Agents need not intend all the expected side effects of their intentions.

If I believe that going to the dentist involves pain, I may believe that going to the dentist involves pain, but this does not imply that I intend to suffer pain! And I may also intend to go to the dentist — but this problem is known as the *dentist problem*.

Agents need not intend all the expected side effects of their intentions.

\[ \phi \Rightarrow \psi \]

\[ \phi \not\Rightarrow \psi \]
Each world is infinitely long linear sequence of states.

Semantics are possible worlds.

- Action $a$ has just happened
  $\text{Done} a$
- Action $a$ happens next
  $\text{Happens} a$
- $x$ has goal of $\phi$
  $\text{Goal} x \phi$
- $x$ believes $\phi$
  $\text{Bel} x \phi$

Following major constructs:

- Cohen-Levesque use a multi-modal logic with the
Each agent allocated:

- **Belief accessibility relation** — $B$

  For every agent/time pair, gives a set of belief accessible worlds.

- **Goal accessibility relation** — $G$

  Euclidean, serial, transitive — gives goal logic $KD45$.

  Serial — gives goal logic $KD$. Serial, serial, transitive — gives belief logic $KD45$.

- **Belief accessibility relation** — $B$

  Each agent allocated:
An Introduction to Multiagent Systems

Chapter 17

A constraint: $G \subseteq B$.

Gives the following inter-modal validity:

Another constraint:

- A realism property — agents accept the inevitable.

- A constraint: $C \subseteq B$.

C&L claim this assumption captures following properties:

- Agents do not indefinitely defer working on goals.
- Agents do not persist with goals forever.

- $\phi \models (\text{Goal}_i \phi) \Rightarrow \Diamond \neg (\text{Goal}_i \phi)$

- $\phi \mathrel{\models} (\text{Bel}_i \phi) \iff \neg (\text{Goal}_i \phi)$

- Properties:

  - $\models \phi \Rightarrow \models \Diamond \neg (\text{Goal}_i \phi)$
  
  - $\models (\text{Goal}_i \phi) \Rightarrow \Diamond \neg (\text{Goal}_i \phi)$
Add some operators for describing the structure of event sequences $\alpha$; $\alpha'\alpha$ followed by $\alpha'$?

'\text{test action}' $\alpha\alpha$

Also add some operators of temporal logic, "\text{always}, and "\text{sometime} (same time) can be defined as (always), and "\text{sometime} (same time) can be defined as

Later:
\\[$\text{Later}:$\\

Abbreviations, along with a "strict" sometime operator, abbreviations, along with a "strict" sometime operator,

Also add some operators of temporal logic, "\text{always}, and "\text{sometime} (same time) can be defined as

\[
\begin{align*}
\Delta \gamma &\equiv \Delta (\Delta \gamma) \\
\gamma &\equiv \gamma \Box \\
(\gamma \gamma) \cdot [\gamma \gamma] &\equiv \gamma \Box
\end{align*}
\]

event sequences

event sequences

And in some operators for describing the structure of

And in some operators for describing the structure of

\[d \hbar \lor d \ll = (d \hbar) \]

\[\Delta \gamma \equiv \Delta \Delta \gamma \]

\[\gamma \hbar \equiv \gamma \Box \]

\[\Delta (\gamma \gamma) \cdot [\gamma \gamma] \equiv \gamma \Box \]
Finally, a temporal precedence operator, (Before).

First major derived construct is a persistent goal.

\[
\begin{align*}
\neg (d \text{ Goal}) & \iff (d \text{ Control} \land (d \text{ Goal}))
\end{align*}
\]
So, an agent has a persistent goal if:

1. It has a goal that eventually becomes true; and
2. Before it drops the goal, one of the following conditions must hold:
   - the agent believes the goal will never be satisfied;
   - the agent believes the goal has been satisfied; or
   - the agent believes the goal eventually becomes true, and $d$ believes that $d$ is not currently true.

So, an agent has a persistent goal if:

[URL: http://www.csc.liv.ac.uk/~mjw/pubs/imas/]

Chapter 17
An Introduction to Multiagent Systems 2e
Chapter 17

An Introduction to Multiagent Systems

Next, intention:

Main point: avoids over commitment.

For intention, C&L discuss how this definition satisfies desiderata and then done. α. α

So, an agent has an intention to do α if it has a persistent goal to have believed it was about to do α,

\[
\forall x \left( \text{Done } x \land \text{Bel}_x (\text{Happens } \alpha) \land \neg \text{Goal}_x (p) \right) \rightarrow \exists \alpha \exists x \left( \text{Intend}_x (\alpha) \right)
\]

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Adaptation of definition allows for relativised intentions. Example:

I have an intention to prepare slides for the tutorial. If I ever come to believe that I will not be paid, the intention evaporates. Relative to the belief that I will be paid for tutoring, I have an intention to prepare slides for the tutorial.
Critique of C&L theory of intention (Singh, 1992):

- disallows multiple intentions.
- does — fully elaborated intentions;
- requires that agents know what they are about to composite actions;
- does not adequately represent intentions to do "competence";
- does not capture and adequate notion of

Critique of C&L theory of intention (Singh, 1992):
C&L use their dynamic logic-style formalism for representing these actions. The key observation is that illocutionary acts are complex events of several speech acts. C&L used their theory of intention to develop a theory of semantics for speech acts.

We will look at request types (of actions).
• First, define alternating belief: \((d \land x \forall u \text{ AltBel}) \cdot u \equiv (d \land x \forall \text{ Bel})\)

And the related concept of mutual belief:

\(\forall n \cdot (\text{AltBel}^n) \equiv (d \land x \forall \text{ Bel})\)

\[\text{First, define alternating belief.}\]
An attempt is defined as a complex action expression.

\[
\{ \text{Attempt } x \text{ epq} \} \equiv \begin{array}{c}
\forall e' \exists e \forall e'' \exists e'''
(\text{Goal } x e'')
\land
(\text{Happens } x e'')
\land
(\text{Bel } x e''')
\land
(\text{Intend } x e''')
\end{array}
\]

(Hence the use of curly brackets, to distinguish from predicate or modal operator.)

An attempt is defined as a complex action expression.
An attempt is a complex action that agents perform when they do something (\(d\)) desiring to bring about some effect (\(e\)) but with intent to produce at least some result (\(b\)) that represents ultimate goal that agent is aiming for by doing \(e\).

Here:

- \(p\) represents ultimate goal that agent is aiming for by doing \(e\).
- \(q\) represents what it takes to at least make an "honest effort" to achieve \(p\).
- \(b\) represents proposition what it takes to at least make an "honest effort" to achieve \(p\).
- \(d\) represents result that agent is aiming for.

In English:
Definition of helpfulness needed:

\[ \text{Helpful}(x, y) \equiv \forall e \cdot \left[ \left( \text{Goal}(x) \land \Diamond (\text{Done}(x)) \right) \leftrightarrow \left[ \left( \Box \neg \text{Goal}(x) \right) \land \left( \forall e \cdot \left( \neg \Diamond \neg \text{Goal}(x) \right) \right) \right] \right] \]

An Introduction to Multiagent Systems 2e
In English:

"Consider an agent \( x \) to be helpful to another agent \( y \) if, for any action \( e \) he adopts the other agent’s goal that he eventually do that action, whenever such a goal would not conflict with his own."
Definition of requests:

\[ \{ \text{Request} \spkr \text{addr} \alpha \} \]

\[ \hat{=} \{ \text{Attempt} \spkr \phi (M - \text{Bel} \text{addr} \alpha (\text{Goal} \spkr \phi (\text{Done} \text{addr} \alpha)))) \]

\[ \text{Intend addr} \alpha \]

\[ \text{Done addr} \alpha \]

\[ \text{Helpful addr} \spkr \alpha \]

\[ \forall \text{ addr } \alpha \]

Where \( \phi \) is

\[ \{ \text{Request} \spkr \text{addr} \alpha \} \]

\[ \begin{align*}
\phi \in \text{spkr} e & \text{ of } \text{spkr} e \in \text{spkr} e \\
\text{Attempt} \spkr \text{addr} \alpha & \text{ of } (\text{spkr} e, \text{addr} \alpha) \\
\text{Bel} \text{addr} \alpha & \text{ of } (\text{spkr} e, \text{addr} \alpha) \\
\end{align*} \]
In English:

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 α (relative to the speaker still having that goal), and 2) the addressee actually eventually does α, or at least brings about a state where the addressee believes α. This is mutually believed that it wants the ideal situation.

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
By this definition, there is no primitive request act.
involved with achieving this desire.

The theory intends to explain how an agent can start

The theory draws on work such as C&L’s model of

precisely, cooperative problem solving).

We now move on to a theory of cooperation (or more

10 A Theory of Cooperation

An Introduction to Multiagent Systems 2e
We formalise our theory by expressing it in a quantified multi-modal logic.

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

Another formal framework
Formal semantics in the paper
1. Recognition.

CPS begins when some agent recognises the potential for cooperative action. This may happen because an agent has a goal that it is unable to achieve in isolation, or because an agent prefers assistance, or because an agent has a goal that it is unable to achieve in isolation, or because the agent

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

The agent that recognised the potential for cooperative action at stage (1) solicits assistance.

2. Team formation.

If team formation successful, then it will end with a group having a joint commitment to collective action.
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.
cooperative action can achieve it. The agent is unable to achieve its goal in isolation, due to a lack of resources, but believes that cooperative action can achieve it.

- Recognition may occur for several reasons:
  - Recognition begins when some agent in a has a goal, and recognizes the potential for cooperative action with respect to that goal.
  - Recognition typically begins when some agent in a has a goal, and recognizes that the agent is unable to achieve its goal in isolation, but believes that cooperative action can achieve it.

12.1 Recognition
An agent may have the resources to achieve the goal, but does not want to use them. It may believe that in working alone on this particular problem, it will clobber one of its other goals, or it may believe that a cooperative solution will be better.
• Formally...

$$\phi[x] = (\text{Achives } \phi) \land \exists g \cdot (\text{Bel } i (J \land \text{Can } g \phi)) \land \left[ \neg (\text{Can } i \phi) \lor (\text{Bel } i \forall \alpha \cdot (\text{Agt } \alpha i) \land \text{Achieves } \alpha \phi) \Rightarrow (\text{Goal } i (\text{Doesnt } \alpha)) \right]$$

\[\text{Note:} \]

- Can is essentially Moore's
- J - Can is a generalization of Moore's
- Can is essentially Moore's

\[\text{Formally...} \]
12.2 Team Formation

Having identified the potential for cooperative action, a rational agent will attempt it.

Note that an agent cannot guarantee that it will be successful in forming a team; it can only attempt it.

If the agent is successful, then it will have brought about a mental state wherein the group has a joint commitment to collective action.

If the agent believes it can achieve the goal, it solicits assistance from some group of agents that will respect its goals, a rational agent will.

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

Note that:

\[
\pre(Team) \phi \Rightarrow \text{pre} \left( \phi \phi \exists \text{Goal} \right) \vee \left( \phi \exists \text{Can} \right) \neg \exists \text{Bel} \left( \phi \phi \exists \text{Goal} \right)
\]

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

An Introduction to Multiagent Systems 2e

Chapter 17

http://www.csc.liv.ac.uk/~mjw/pubs/imas/71
have occurred. Conditions under which negotiation can be said to
simply offer some observations about the weakest
Unfortunately, negotiation is extremely complex — we
formulation, which involves negotiation.
Hence the next stage in the CPS process: plan
But collective action cannot begin until the group
action.
If team formation is successful, then there will be a

12.3 Plan Formation
Stage to follow.

- If negotiation succeeds, we expect a team action.
  - At least one agent would propose a course of action that it believed would take the collective closer to the goal.
- If negotiations succeed, we expect a team action.
  - 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.

- If negotiations fail, the collective may simply be unable to reach agreement.

- Note that negotiation may fail: the collective may
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.
(\(\exists \alpha \cdot (Happens \{ Attempt g \} \wedge (1 \phi g \text{ Team } g \text{ Know } W) \wedge d)\))

where

\(\{ b d \wedge g \text{ Attempt } \} \wedge \text{ Happens }.cmd \cdot \alpha \text{ Attempts } \)

\(\exists (1 \phi g \text{ Team } g \text{ Know } W) \wedge d)\)

The main assumption is then:

\(\text{PreTeam}(1 \phi g \text{ Team } g \text{ Know } W) \wedge d)\)
12.4 Team Action

Team actions simply involve the team jointly intending to achieve the goal. The formalisation of Team is simple:

\[
\left( \exists \alpha \cdot (\text{Achieves} \, \alpha \land \text{Intend} \, \alpha \land \text{Goal} \, i \, \phi) \right) \land \left( \forall \, (i \phi \exists \, \text{Team}) \right)
\]

http://www.csc.liv.ac.uk/~mjw/pubs/imas/