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.

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 research of agent theories? And tools that we use — literally a formal methods have (arguably) had little impact of general.

Without such a semantics, it is never clear exactly what is happening, or why it works.

In agent-based systems, we have a bag of concepts and tools, but we need theory to reach any kind of profound understanding of these tools.

In agent theory:
- Introduce the Common-Languages Theory of Intention as a case study in agent theory.
- Introduce the Common-Language Theory of Action.
- Discuss Moore’s Theory of Ability.
- Introduce 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 develop semantic theories.
- Begin by answering the question: Why theory?
- 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.

2 Why Theory?
The notion of an agent as an intentional system requires the intentional stance. An agent theorist starts with the (strong) view of agents as intentional systems: one whose simplest consistent description

4 Theories of Attitudes

5 Formalising Attitudes

3 Agents = Intentional Systems

So first, which attitudes?

Consider (Zeus = Jupiter)

Where do theories start from?

Agents = intentional systems. One whose simplest consistent description

Let us go back to the discussion of the notion of intention as mentioned earlier. Consider Janine believes that Cronos is father of Zeus.

Informal translation into first-order logic:

**Believe (Janine, Father (Zeus, Cronos))**

Consider...

• Janine believes that Cronos is father of Zeus.

• Do we formulate attitudes?

But...

Believe (Janine, Father (Zeus, Cronos))

Naive translation into first-order logic:

Janine believes that Cronos is father of Zeus.

Consider...

• So how do we formulate attitudes?

3 Formalising Attitudes

An Introduction to Multiagent Systems

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

2 Two categories:

pro-attitudes

commitment

information attitudes

belief

desire

intention

intellectual notions are referentially opaque.

Consider (Zeus = Jupiter)

allows us to substitute terms with the same denotation:

need to be able to apply, but to formulate:

first-order logic, not a term:

the second argument to the Bel predicate is a formula of

• But...

Believe (Janine, Father (Zeus, Cronos))

Naive translation into first-order logic:

Janine believes that Cronos is father of Zeus.

Consider...

• So how do we formulate attitudes?

5 Formalising Attitudes

An Introduction to Multiagent Systems

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

4 Theories of Attitudes

An Introduction to Multiagent Systems

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

3 Agents = Intentional Systems

An Introduction to Multiagent Systems

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

2 Two categories:

pro-attitudes

commitment

information attitudes

belief

desire

intention

intellectual notions are referentially opaque.

Consider (Zeus = Jupiter)

allows us to substitute terms with the same denotation:

need to be able to apply, but to formulate:

first-order logic, not a term:

the second argument to the Bel predicate is a formula of

• But...

Believe (Janine, Father (Zeus, Cronos))

Naive translation into first-order logic:

Janine believes that Cronos is father of Zeus.

Consider...

• So how do we formulate attitudes?

5 Formalising Attitudes

An Introduction to Multiagent Systems

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

4 Theories of Attitudes

An Introduction to Multiagent Systems

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

3 Agents = Intentional Systems

An Introduction to Multiagent Systems

http://www.csc.liv.ac.uk/~mjw/pubs/imas/
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 (intentional notions refer to equivalences).

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:

- use a modal language, which contains modal operators.
- use a meta-language: a first-order language containing terms that denote formulae of some other object-language.

Wedefine a (propositional) modal logic for knowledge/belief.

Vocabulary:

- primitive propositions
- classical connectives
- modal connective

Syntax:

- any member of \( W \) is allowed.

Example formulae:

\[
\begin{align*}
\langle \phi \rangle w \\
\langle \phi \rangle w \land \langle \psi \rangle w \\
\langle \phi \rangle w \lor \langle \psi \rangle w
\end{align*}
\]

\( \phi \) any member of \( W \)
Semantics are trickier. The idea is that an agent's beliefs can be characterized as a set of possible worlds, in the following way.

Consider an agent playing a card game such as poker, who possessed the ace of spades. How could she deduce what cards were held by her opponents?

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 (for example, any configuration in which she did not possess the ace of spades could be rejected.)

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

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!

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 if the agent was actually in world \( w \), then as far as it was concerned, it might be in world \( w' \).

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

- Something true in all our agent's possibilities is believed by the agent.
- The following axiom schema is valid:
  \[ K \phi \Rightarrow \phi \]
  \[ K \phi \Rightarrow K K \phi \]
  The following axiom schema is valid:
  \[ (K \phi \land K \psi) \Rightarrow (K \phi \rightarrow K \psi) \]
  Thus agent's knowledge is closed under logical consequence. This is not a desirable property!
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 various axioms; there are lots of these, but the most important are:

- Axiom T is the knowledge axiom: if you know \( \phi \), you can't also know \( \neg \phi \).
- Axiom D is the consistency axiom: if you know \( \phi \), you know you don't know \( \neg \phi \).
- Axiom 4 is positive introspection: if you know \( \phi \), you know you know \( \phi \).
- Axiom 5 is negative introspection: you are aware of what you don't know.

All of these (KT45) constitute the logical system S5. (Often chosen as a logic of idealised knowledge.)

S5 without the T is weak-S5, or KD45. (Often chosen as a logic of idealised belief.)

Moore considered 2 aspects of interaction between knowledge and action:

1. As a result of performing an action, an agent can gain knowledge.
   - Moore's 1977 analysis is best-known in this area.
   - Culminated in defining ability: what it means to do bring something about.

2. In order to perform some actions, an agent needs knowledge: these are knowledge pre-conditions.
   - Formal tools: a modal logic with Kripke semantics + dynamic logic-style representation for action.
   - But showed how Kripke semantics could be axiomatized in a first-order meta-language.
   - Modal formulas then translated to meta-language using axiomatization.
   - Modal theorem proving reduces to meta-language theorem proving.
Axiomatising standard logical connectives:

- First conjunct says the action is possible:
  \[ \phi \iff (\mu w A \lor (\mu w A) w) \]

- Second says that a necessary consequence of performing action \( \mu w A \) will be true:

  \[ (\phi \iff (\mu w A)) \iff (\phi \mu w A) \]


These axioms ensure that \( \phi \) is an equivalence relation:

\[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Reflexive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Transitive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Euclidean:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

These axioms ensure that \( \phi \) is an equivalence relation:

\[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Reflexive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Transitive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Euclidean:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

These axioms ensure that \( \phi \) is an equivalence relation:

\[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Reflexive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Transitive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Euclidean:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

These axioms ensure that \( \phi \) is an equivalence relation:

\[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Reflexive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Transitive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Euclidean:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

These axioms ensure that \( \phi \) is an equivalence relation:

\[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Reflexive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Transitive:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]

- Euclidean:
  \[ (\mu w A) \equiv (\mu w A) \equiv (\mu w A) \]
Now we can define ability via modal operators:

\[ \text{If is my intention to prepare my slides.} \]

Here we mean intention as in...

The theory of intention developed by Cohen & Levesque:

1. We need a set of connectives for talking about an agent's pro-attitudes and plans as well.
2. Agent needs to achieve a rational balance between its attitudes:
   - Should not be over-committed.
   - Should not be under-committed.
3. Agent knows the identity of the action.
   - We have one aspect of an agent's believes/past belief.
4. Agent performs an action to find out how to achieve goal.
   - We can weaken the definition to capture the case where an

8 Intention

No, it's not a fixed point.

A circular definition?

\[ ( ((\phi \land \text{know} \psi) \land \text{perform} \phi) \land \text{true} ) \land \frac{W \land ( ((\phi \land \text{true}) \land \text{perform} \phi) \land \frac{m \land \phi) \land \frac{A} {A} }{A} \]

Critique of Moore's formism:

1. Translating modal language into a first-order language will be more efficient.
2. Translating modal language into a first-order language is inefficient.

Critique of Moore's formism:

1. Translating modal language into a first-order language is inefficient.
2. Formulae resulting from the translation process are complicated and unintuitive.
3. Logical omniscience (and hence sense) is lost.
4. Original structure (and hence sense) is lost.
5. Moore's formalism is somewhat vacuous.
6. Hard-wired modal theorem provers will be more efficient.
7. Definition of ability is somewhat vacuous.
8. Definition of ability is somewhat vacuous.
9. Definition of ability is somewhat vacuous.
8.1 What is intention?

Twosorts:
– presentdirectedattitudetoanactionfunctioncausallyinproducingbehaviour.
– futuredirectedattitudetoapropositionservetocoordinatefutureactivity.

Wearehereconcernedwithfuturedirectedintentions.

FollowingBratman(1987)Cohen-Levesqueidentifysevenpropertiesthatmustbesatisfiedbyintention:
1. Intentionsposeproblemsforagents,whoneedtodetermine
   waysofachievingthem.
2. Intentionsposeproblemsforagents,whoneedto
determine
   propertiesthatmustbesatisfiedbyintention:
   http://www.csc.liv.ac.uk/~mjw/pubs/imas/
Cohen-Levesque use multi-modal logic with the following major constructs:

- $x$ believes
- $x$ has goal of
- action happens next
- action has just happened
- goal of
- $x$ believes

Semantics are possible worlds.

- Each world is a linear sequence of states.
- Each world is infinitely long.
- Each agent is allocated a set of goal accessible worlds.
- Each agent has a set of belief accessible worlds.
- Each agent accepts the inevitable.

C&L claim this assumption captures the following properties:

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

Another constraint:

A realism property — agents accept the inevitable.

- A constraint: $G \subseteq B$.

Add in some operators for describing the structure of event sequences.

Also add some operators of temporal logic. 

- $\lozenge$ followed by
- $\diamond$ test action
- $\lozenge$ always

Coh-Levesque use a multi-modal logic with the following operators:

- Goal accessibility relation $G$
- Belief accessibility relation $B$
- Euclidean, serial, transitive gives belief logic $K D 4 5$. For every agent/time pair, gives a set of goal accessible worlds.
- $\Box \forall y \exists x (y < x)$
- $\Diamond$ always
- $\lozenge$ test action
- $\lozenge$ always
- $\Box$ followed by
- $\lozenge$ test action
- $\lozenge$ always

\[ \phi \land \Box \phi \equiv (\phi \land \Box \phi) \]
\[ \Box \phi \equiv \Box \phi \]

$\Box \forall y \exists x (y < x)$

Each world is infinitely long.
Finally, at a temporal precedence operator, $pq$.

First major derived construct is a persistent goal.

1. The agent believes the goal will never be satisfied.
2. The agent believes the goal has been satisfied.
3. The agent's belief that the goal will never be satisfied becomes true, and believes $\neg p$
4. So an agent has a persistent goal of $p$: $\neg p$
5. First major derived construct is a persistent goal.

Next, intention:

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

Main point: avoids ever commitment.

Adaptation of definition allows for relativised intentions. Example:

Critique of C&L theory of intention (Singh, 1992):

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

So, an agent has an intention to do $a$ if $a$ is a persistent goal.
C\&L used their theory of intention to develop a theory of speech acts.

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

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

We will look at request.

First, define alternating belief.

We will look at request.

These actions.

C\&L use their dynamic logic-style formalism for representing illocutionary acts as complex event types (cf. speech acts).

And the related concept of mutual belief.

\[ (d \land u) \land (d \land \neg \perp) \]

And the related concept of mutual belief.

\[ \text{Semantics for Speech Acts} \]
Definition of helpfulness needed:

\[ \text{helpful}(x, y) \equiv \exists \phi (\text{Done} x (\phi) \land (\text{Bel} y (\text{Goal} x (\phi) \land \neg (\text{Done} y (\phi)))) \land \neg (\text{Conflict} x (\phi) \land \text{inclin} x (\phi))) \]

In English:

“Consider an agent \( y \) to be helpful to another agent \( x \) if, for any action \( \alpha \) that \( x \) adopts, the other agent's goal that he eventually does that action, whenever such a goal would not conflict with his own.”

Definition of requests:

\[ \text{request} (\text{spkr}, \text{addr}, \alpha) \equiv \{ \text{Attempt} \text{spkr} \varepsilon (\text{Goal} \text{spkr} (\alpha)) \land \neg (\text{Bel} \text{addr} \varepsilon (\text{Done} \text{addr} (\alpha))) \land ((\text{Helpful} \text{spkr} (\alpha)) \land \neg (\text{Conflict} \text{spkr} (\alpha))) \}

In English:

“\( \alpha \) is viewed as having performed a request if he executes any sequence of actions that produces the needed effects.”

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.”

A request is an attempt on the part of \( \text{spkr} \), by doing \( \alpha \), to bring about a state where, ideally, 1) \( \text{addr} \) intends \( \alpha \) (relative to the \( \text{spkr} \) still having that goal, and \( \text{addr} \) still being helpfully inclined to \( \text{spkr} \)), and 2) \( \text{addr} \) actually eventually does \( \alpha \), or at least brings about a state where \( \text{addr} \) believes it is mutually believed that it wants the ideal situation.”

In English:

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

We now move on to a theory of cooperation (or more precisely, cooperation in problem solving).

This theory aims to explain how an agent can start with an initial commitment to collective action. It is based on work such as Clark's model of intention, and our formal semantics for speech acts.

The theory involves explaining how an agent can start with an initial commitment to a group, and then build blocks.

This theory grew out of work such as Clark's model of intention, and the formal semantics for speech acts.

The theory is intended to explain how an agent can start with an initial commitment to collective action.

We now move on to a theory of cooperation (or more precisely, cooperative problem solving).

11 Another (Formal) Framework

Formal semantics in the paper

- Desires and beliefs (rational choice in branching time structure associated with agents).
- Actions (transitions in branching time structure associated with agents).
- Dynamic logic style action constructors:
  - Goals
  - Beliefs
- Multi-modal logic.

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

3. Plan formation.

The newly agreed plan of joint action is executed by the agents.

4. Team action.

Achieve the desired goal.

The agents attempt to negotiate a joint plan that they believe will achieve the desired goal.

This theory draws on works such as C&L's model of intention, and their semantics for speech acts.

It uses connectives such as intend as the building blocks of cooperative problem solving.

10 A Theory of Cooperation

Formalisation in the paper

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

Formal semantics in the paper

With agents:

- Actions (transitions in branching time structure associated with agents).
- Dynamic logic style action constructors:
  - Goals
  - Beliefs
- Multi-modal logic.

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

1. Recognition.

CPS begins when some agent recognises the potential for cooperation. Maybe happens because an agent has a goal that it is unable to achieve in isolation, or because the agent prefers assistance.

This theory grew out of work such as Clark's model of intention, and the formal semantics for speech acts.

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

Team formation.

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

4. Team formation.

A(nother) Formal Framework

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

- Beliefs;
- Goals;
- Dynamic logic style action constructors;
- Groups (sets of agents) as terms in the language — set quantifiers (branching time);
- Predefined (branching time).

The newly agreed plan of joint action is executed by the agents.

1. Recognition.

CPS begins when some agent recognises the potential for cooperation.

Recognition.

2. Team formation.

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

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.

Team formation.

The newly agreed plan of joint action is executed by the agents.

We move on to a theory of cooperation (or more precisely, cooperative problem solving).

This theory is intended to explain how an agent can start with an initial commitment to collective action.
12.1 Recognition

CPStypicallybeginswhensomeagentinahasagoal, and recognizes the potential for cooperative action with respect to that goal.

Recognitionmayoccurforseveralreasons:
– Theagentisunabletoachieveitisolation, duetoalackofresources, butbelievesthatcooperativeactioncan achieveit.
– Anagentmayhaveadditionalresources, notused. Itmaybelievethatinworkingaloneonthisparticularproblem, itwill achieveitisolation, due to a lack of resources, butbelieves that cooperative action can achieveitisolation. That goal achievement is especially beneficial when someagentinahasagoodgoal, and achievesit.
The main assumption concerning team formation can now be stated.

\[ \forall i: (\text{Bel}_i(\text{Potential} - \text{for} - \text{Coop}(\phi)) \Rightarrow (\text{Gain}_j(\phi) \land \text{Happens}(\text{Attempt}_i \land p \land q))) \]

where

\[ p = (\text{Pre}_\text{Team}(\phi \land \phi)) \]

\[ q = ((\text{M} - \text{Bel}_i(\text{Goal}_i(\phi)) \land (\text{Bel}_i(J - \text{Can}_g(\phi))))) \]

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:

\[ p = (\text{M} - \text{Know}_g(\text{Team}_g(\phi))) \]

\[ q = ((\forall \alpha: (\text{M} - \text{Bel}_i(\text{Gain}_i(\phi)) \land (\text{Att}_\text{Attempt}_i \land g \land p \land q))) \]

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

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 negotiation succeeds, we expect a team action stage to follow.
Team actions simply involve the team jointly intending to achieve the goal.

The formalisation of Team is simple:

\[
\begin{align*}
&\left( \phi \land \text{achieves } \phi \right) \\
&\lor \left( \phi \land \text{achieves } \phi \land \text{achieves } \phi \right) \\
&\equiv \left( \frac{\phi}{\text{achieves } \phi} \right)
\end{align*}
\]