A4M33MAS - Multiagent Systems

Belief-Desire-Intention (BDI) Architecture & Social Commitments

Michal Pechoucek & Michal Jakob

Department of Computer Science
Czech Technical University in Prague

OTEVŘENÁ INFORMATIKA

In parts based on Michal Pechoucek: Multi-Agent Systems, Lecture course at State University of New York, University at Binghamton

Towards Architectures for IA

- Reactive Architectures
- Deliberative Architectures

Models of Practical Reasoning: BDI

process of figuring out what to do -- practical reasoning is a matter of weighing conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires/values/cares about and what the agent believes (Bratman)

- computational model of human decision process oriented towards an action, based on models of existing mental models of the agents
- human practical reasoning consists of two activities:
  - deliberation: deciding what state of affairs we want to achieve and
  - means-ends reasoning (planning): deciding how to achieve these states
- the outputs of deliberation process are intentions

BDI Architecture

- BELIEFS
  - collection of information that the agents has about its the status of the environment, peer agents, self
- DESIRES
  - set of long term goals the agent wants to achieve
- INTENTIONS
  - agents immediate commitment to executing an action, either high-level or low level (depends on agents planning horizon)

- BDI architecture connects: (i) reactive (ii) planning & (iii) logical representation. BDI architecture does not count on theorem proving

$$\text{if } \varphi \in \mathcal{L}_{agent} \text{ then } \varphi, (\text{Bel } A\ \varphi), (\text{Des } A\ \varphi), (\text{Int } A\ \varphi) \in \mathcal{L}_{bdi}$$

BDI Inference Algorithm

- Basic algorithm:

1. initial beliefs → Bel
2. while true do
3. Read(get_next_percept) → in
4. Belief-revision(Bel, in) → Bel
5. Deliberate(Bel, Des) → Int
6. Plan(Bel, Int) → π
7. Execute(π)
8. end while

BDI Modal Properties

- BELIEFS
  - KD45 system, modal logic where the B relation is serial, transitive and euclidean: satisfies K axioms, positive introspection axiom (4 axiom), negative introspection axiom (5 axiom), beliefs consistency axiom (D axiom).

BDI Modal Properties

- BELIEFS
  - KD45 system, modal logic where the B relation is serial, transitive and euclidean: satisfies K axioms, positive introspection axiom (4 axiom), negative introspection axiom (5 axiom), beliefs consistency axiom (D axiom).

- DESIRES
  - KD system, modal logic requiring desired goals not to contradict (D axiom).

$$(\text{Des } A\ \varphi) \rightarrow \neg(\text{Des } \neg\ \varphi)$$

BDI Modal Properties

- BELIEFS
  - KD45 system, modal logic where the B relation is serial, transitive and euclidean: satisfies K axioms, positive introspection axiom (4 axiom), negative introspection axiom (5 axiom), beliefs consistency axiom (D axiom).

- DESIRES
  - KD system, modal logic requiring desired goals not to contradict (D axiom).

$$(\text{Des } A\ \varphi) \rightarrow \neg(\text{Des} \neg\ \varphi)$$

- INTENTIONS
  - KD system, modal logic requiring intentions not to contradict (D axiom).

$$(\text{Int } A\ \varphi) \rightarrow \neg(\text{Int } A \neg\ \varphi)$$

Properties of Intentions

- Intention persistency:
  - agents track the success of their intentions, and are inclined to try again if their attempts fail

$$(\mathsf{Int}\ A\ \varphi) \curvearrowleft \varphi$$

- Intention satisfiability:
  - agents believe their intentions are possible; that is, they believe there is at least some way that the intentions could be brought about.

$$(\mathsf{Int}\ A\ \varphi) \Rightarrow \mathsf{EF}\varphi$$

Properties of Intentions

- Intention-belief inconsistency:
  – agents do not believe they will not bring about their intentions; it would be irrational of agents to adopt an intention if believed was not possible

$$(\mathsf{Int}\ A\ \varphi) \wedge (\mathsf{Bel}\ A \neg \mathsf{EF}\varphi)$$

- Intention-belief incompleteness:
  – agent do not believe that their intention is possible to be achieved, may be understood as rational behavior

$$(\mathsf{Int}\ A\ \varphi) \wedge (\neg \mathsf{Bel}\ A \mathsf{EF}\varphi)$$

  – agents admit that their intentions may not be implemented.

$$(\mathsf{Int}\ A\ \varphi) \wedge (\mathsf{Bel}\ A\ \mathsf{EF}\neg\varphi)$$

Properties of Intentions

- Intention side-effects:
  - Agents need not intend all the expected side effects of their intentions. Intentions are not closed under implication.

$$(\text{Bel}\ A\ \psi \Rightarrow \varphi) \wedge (\text{Int}\ A\ \psi) \wedge \neg(\text{Int}\ A\ \varphi)$$

  - * *is thus classified as fully rational behavior*

  - <u>Example:</u> I may believe that going to the dentist involves pain, and I may also intend to go to the dentist - but this does not imply that I intend to suffer pain!

Rationality of Inevitables & Options

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Rationality of Inevitables & Options

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Rationality of Inevitables & Options

01

1. **inevitables:**

$$(\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{AG}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{AG}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{AG}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{AG}\varphi)$$

2. **options:**

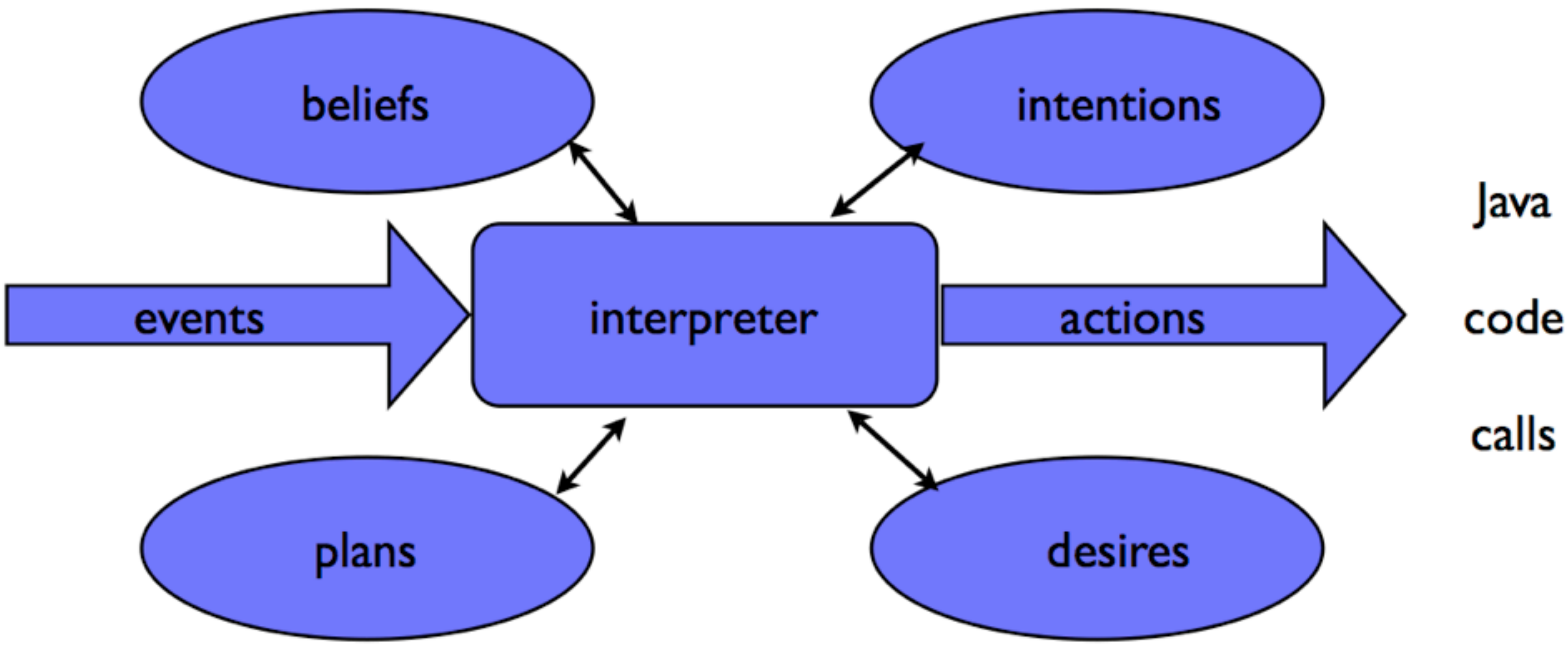
$$(\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Des}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Int}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$
$$(\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Des}\ A\ \mathsf{EF}\varphi) \qquad (\mathsf{Bel}\ A\ \mathsf{EF}\varphi) \Rightarrow (\mathsf{Int}\ A\ \mathsf{EF}\varphi)$$

Example: Model Checking AgentSpeak

- AgentSpeak(L) is a BDI programming language introduced by Rao.
- A simple but powerful programming language for building rational agents. Based on Prolog.
- Jason:
  - implementation of AgentSpeak in Java
  - A development environment for AgentSpeak systems

Example: Model Checking AgentSpeak

- AgentSpeak(L) is a BDI programming language introduced by Rao.
- A simple but powerful programming language for building rational agents. Based on Prolog.



AgentSpeak Control Loop

1. agent receives events, which are either
   - external (from the environment, from perceptual data)
   - internally generated
2. tries to handle events by looking for plans that match the event and lead to the goal → desires (options)
3. chooses one plan from its desires to execute: becomes committed to it → intention
4. as it executes a plan may generate new events that require handling

AgentSpeak: BDI primitives

- Beliefs: Symbolically represented (ground atoms or FOL formulas)

$$ag = \langle bs, ps \rangle: \ \mathrm{BEL}_{\langle ag, C \rangle}(\varphi) \equiv \varphi \in bs.$$

AgentSpeak: BDI primitives

- Beliefs: Symbolically represented (ground atoms or FOL formulas)

$$ag = \langle bs, ps \rangle: \;\; \mathrm{BEL}_{\langle ag, C \rangle}(\varphi) \equiv \varphi \in bs.$$

agent beliefs plans context

$(\mathsf{Bel}\ A\ \varphi)$

AgentSpeak: BDI primitives

- Beliefs: Symbolically represented (ground atoms or FOL formulas)

$$ag = \langle bs, ps \rangle: \ \mathrm{BEL}_{\langle ag, C \rangle}(\varphi) \equiv \varphi \in bs.$$

- <u>Manipulating beliefs:</u>

  +B adding new belief

  -B dropping belief

AgentSpeak: BDI primitives

- Beliefs: Symbolically represented (ground atoms or FOL formulas)

$$ag = \langle bs, ps \rangle: \ \mathrm{BEL}_{\langle ag, C \rangle}(\varphi) \equiv \varphi \in bs.$$

- Manipulating beliefs:

  +B adding new belief

  -B dropping belief

- Manipulating goals/intentions:

  +!D adding new desire

  -!D dropping desire

- Plans:

```
triggerCondition :
                context <-
                body.
```

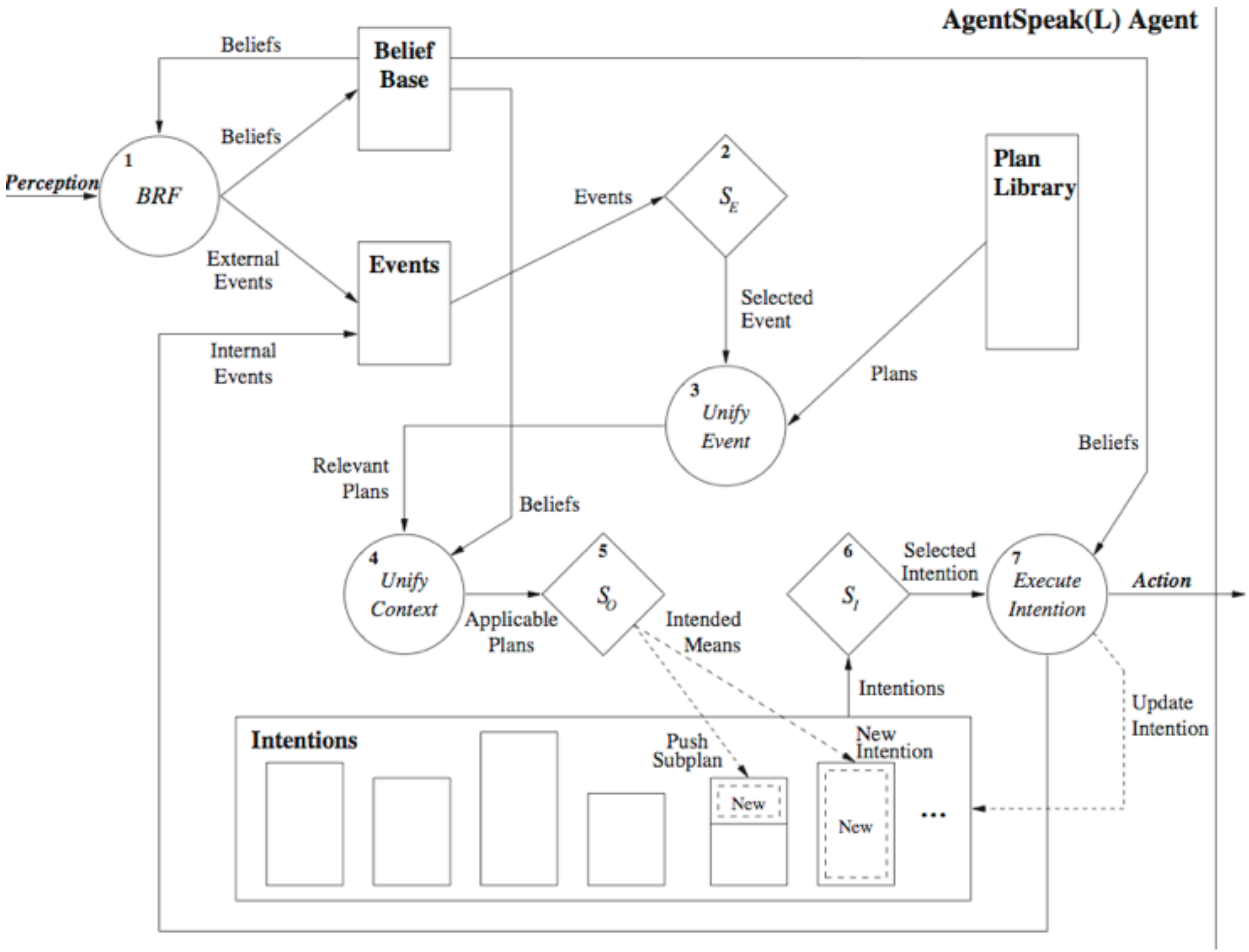
AgentSpeak: Example


```
+green_patch(Rock)
    : not battery_charge(low)
      <- ?location(Rock,Coordinates);
         !traverse(Coordinates);
         !examine(Rock).

+!traverse(Coords)
    : safe_path(Coords)
      <- move_towards(Coords).

+!traverse(Coords) :
    : not safe_path(Coords)
      <- ...
```

AgentSpeak Reasoning Lifecycle



AgentSpeak: Example

Agent1

```
+auction(N) : true
   <- place_bid(N,6).
```

Agent2

```
myself(ag2).
bid(ag2,4).
ally(ag3).

+auction(N) : myself(I) & ally(A) & not(alliance(A,I))
   <- ?bid(I,B); place_bid(N,B).

+auction(N) : myself(I) & ally(A) & (alliance(A,I))
   <- place_bid(N,0).

+alliance(A,I) : myself(I) & ally(A)
   <- ?bid(I,B);
      .send(A,tell,bid(I,B));
      .send(A,tell,alliance(A,I)).
```

AgentSpeak: Example

Agent3

```
myself(ag3).
bid(ag3,3).
ally(ag2).
threshold(3).
```

```
+auction(N) : threshold(T) & .gte(T,N)
   <- !bid_normally(N).

+auction(N) : myself(I) & winner(I)
              & ally(A) & not(alliance(I,A))
   <- !bid_normally(N).

+auction(N) : myself(I) & not(winner(I))
              & ally(A) & not(alliance(I,A))
   <- !alliance(I,A);
      !bid_normally(N).

+auction(N) : alliance(I,A)
   <- ?bid(I,B); ?bid(A,C);
      .plus(B,C,D); place_bid(N,D).

+!bid_normally(N) : true
   <- ?bid(I,B); place_bid(N,B).

+!alliance(I,A) : not(alliance(I,A))
   <- .send(A,tell,alliance(I,A)).
```

AgentSpeak: Example

$$\Box\left(\begin{array}{l}\neg(\text{Bel } \mathtt{ag3}\ \mathtt{winner}(\mathtt{ag3}))\wedge\\ (\text{Des } \mathtt{ag3}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\Rightarrow\\ \Diamond(\text{Int } \mathtt{ag3}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\end{array}\right)$$

$$\Diamond\left(\begin{array}{l}(\text{Bel } \mathtt{ag2}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\wedge\\ (\text{Bel } \mathtt{ag3}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\end{array}\right)$$

$$\Box\left(\begin{array}{l}(\text{Bel } \mathtt{ag2}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\wedge\\ (\text{Bel } \mathtt{ag3}\ \mathtt{alliance}(\mathtt{ag3},\mathtt{ag2}))\Rightarrow\\ \Diamond\Box\mathtt{winner}(\mathtt{ag3})\end{array}\right)$$

Social Commitments

Agents Individual/Social Commitments

- Commitments: knowledge structure, declarative programming concept based on intentions (intentions are special kinds of comms).
  - specify relationships among different intentional states of the agents
  - specify social relations among agents, based on their comms to joint actions

The commitment is an agent's state of 'the mind' where it commits to adopting the single specific intention or a longer term desire.

- We distinguish between:
  - specific, commonly used commitments general commitments
  - individual commitments social commitments

Individual Commitments

- A can get committed to its intention φ in several different ways:$\varphi\varphi$
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft (\mathsf{Bel}\ A\ \varphi))$$

Individual Commitments

- A can get committed to its intention φ in several different ways:$\varphi\varphi$
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft (\mathsf{Bel}\ A\ \varphi))$$

  - single-minded commitment: besides above it intends the intention until it believes that it is no longer possible to achieve the goal

$$(\mathsf{Commit}\ A\ \varphi) \equiv \mathsf{AG}((\mathsf{Int}\ A\ \varphi) \curvearrowleft ((\mathsf{Bel}\ A\ \varphi) \vee (\mathsf{Bel}\ A\ \neg \mathsf{EF}\varphi))$$

Individual Commitments

- A can get committed to its intention φ in several different ways:$\varphi\varphi$
  - blind commitment: also referred to as fanatical commitment, the agent is intending the intention until it believes that it has been achieved (persistent intention)

$$(\text{Commit } A\ \varphi) \equiv \text{AG}((\text{Int } A\ \varphi) \curvearrowleft (\text{Bel } A\ \varphi))$$

  - single-minded commitment: besides above it intends the intention until it believes that it is no longer possible to achieve the goal

$$(\text{Commit } A\ \varphi) \equiv \text{AG}((\text{Int } A\ \varphi) \curvearrowleft ((\text{Bel } A\ \varphi) \vee (\text{Bel } A\ \neg\text{EF}\varphi))$$

  - open-minded commitment: besides above it intends the intention as long as it is sure that the intention is achievable

$$(\text{Commit } A\ \varphi) \equiv \text{AG}((\text{Int } A\ \varphi) \curvearrowleft ((\text{Bel } A\ \varphi) \vee \neg(\text{Bel } A\ \text{EF}\varphi))$$

General Commitments

- Commitment is defined as $(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda)$, where
- Convention is defined as $\lambda = \{\langle\rho_k, \gamma_k\rangle\}_{k\in\{1,\ldots,l\}}$
  - provided $\curvearrowleft$ stands for *until*, A stands for *always in the future*, Int is *agent's intention* and Bel is *agent's belief* then for $\lambda = \langle\rho, \gamma\rangle$ the commitment has the form:

$$(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge \textit{decommitment\_rule} \qquad) \curvearrowleft \gamma)$$

$$(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge ((\mathsf{Bel}\ A\ \rho) \Rightarrow \mathsf{A}(\mathsf{Int}\ A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

General Commitments

- Commitment is defined as $(\text{Commit } A \; \varphi \; \psi \; \lambda)$, where
- Convention is defined as $\lambda = \{\langle \rho_k, \gamma_k \rangle\}_{k \in \{1,\ldots,l\}}$
  - provided $\curvearrowleft$ stands for *until*, A stands for *always in the future*, Int is *agent's intention* and Bel is *agent's belief* then for $\lambda = \langle \rho, \gamma \rangle$ the commitment has the form:

$$
\begin{aligned}
(\text{Commit } A \; \varphi \; \psi \; \lambda) \quad \equiv \\
& \psi \wedge \mathsf{A}((\text{Int } A \; \varphi) \wedge \\
& \quad ((\text{Bel } A \; \rho_1) \Rightarrow \mathsf{A}(\text{Int } A \; \gamma_1)) \curvearrowleft \gamma_1) \\
& \quad \ldots \\
& \quad ((\text{Bel } A \; \rho_k) \Rightarrow \mathsf{A}(\text{Int } A \; \gamma_k)) \curvearrowleft \gamma_k) \\
& \curvearrowleft \bigvee_i \gamma_i)
\end{aligned}
$$

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

$$(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge ((\mathsf{Bel}\ A\ \rho) \Rightarrow \mathsf{A}(\mathsf{Int}\ A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)

$$(\mathsf{Commit}\ A\ \varphi\ \psi\ \lambda) \equiv \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge ((\mathsf{Bel}\ A\ \rho) \Rightarrow \mathsf{A}(\mathsf{Int}\ A\ \gamma)) \curvearrowleft \gamma) \curvearrowleft \gamma)$$

$$\begin{aligned}(\mathsf{J\text{-}Commit}\ \Theta\ \varphi\ \psi\ \lambda) \quad \equiv \quad & \\ & \forall A:\ (A \in \theta) \Rightarrow \\ & \psi \wedge \mathsf{A}((\mathsf{Int}\ A\ \varphi) \wedge \\ & \quad ((\mathsf{Bel}\ A\ \rho) \Rightarrow \mathsf{A}(\mathsf{Int}\ A\ \gamma) \curvearrowleft \gamma) \\ & \curvearrowleft \gamma)\end{aligned}$$

Joint (Social) Commitment

- Form of a commitment that represents how a group of agents is committed to a joint action (goal, intention, ...)
  - for a convention in the form of $\lambda = \{\langle \rho_k, \gamma_k \rangle\}_{k \in \{1,\dots,l\}}$

$$(\text{J-Commit}\ \Theta\ \varphi\ \psi\ \lambda) \equiv \forall A:\ (A \in \theta) \Rightarrow \psi \wedge \mathsf{A}((\chi_1 \wedge \chi_2) \curvearrowleft \chi_3)$$

where

$$\begin{aligned}
\chi_1 &= (\text{Int}\ A\ \varphi) \\
\chi_2 &= ((\text{Bel}\ A\ \rho_1) \Rightarrow \mathsf{A}((\text{Int}\ A\ \gamma_1) \curvearrowleft \gamma_1)) \wedge ((\text{Bel}\ A\ \rho_2) \Rightarrow \\
&\quad \mathsf{A}((\text{Int}\ A\ \gamma_2) \curvearrowleft \gamma_1) \wedge \cdots \wedge ((\text{Bel}\ A\ \rho_n) \Rightarrow \mathsf{A}((\text{Int}\ A\ \gamma_n) \curvearrowleft \gamma_n))) \\
\chi_3 &= \gamma_1 \vee \gamma_1 \vee \cdots \vee \gamma_n
\end{aligned}$$

Blind Social Commitment

- each agent is trying to accomplish the commitment until achieved

$$\lambda_{blind} = \{\langle(\text{Bel } A\ \varphi), (\text{M-Bel } \Theta\ \varphi)\rangle\}$$
$$\psi_{blind} = \neg(\text{Bel } A\ \varphi)$$

$$\begin{aligned}(\text{J-Commit } \Theta\ \varphi\ \psi\ \lambda) \quad \equiv \quad & \forall A : \ (A \in \Theta) \Rightarrow \\ & (\neg(\text{Bel } A\ \varphi) \wedge (\text{A}((\text{Int } A\ \varphi) \wedge \\ & \quad ((\text{Bel } A\ \varphi) \Rightarrow \text{A}((\text{Int } A\ (\text{M-Bel } \Theta\ \varphi)) \\ & \quad \curvearrowleft (\text{M-Bel } \Theta\ \varphi)))) \\ & \curvearrowleft (\text{M-Bel } \Theta\ \varphi))).\end{aligned}$$

Minimal Social Commitment

- minimal social commitment, also related to as joint persistent goal:
  - initially agents do not believe that goal is true but it is possible
  - every agent has the goal until *termination condition* is true
  - until termination: if agent beliefs that the goal is either *true or impossible* than it will want the goal that it becomes a mutually believed, but keep committed
  - the termination condition is that it is *mutually believed* either goal is true or impossible to be true.

$$\psi_{soc} \quad = \quad \neg(\mathsf{Bel}\ A\ \varphi) \wedge (\mathsf{Bel}\ A\ \mathsf{EF}\varphi)$$

$$\lambda_{soc} \quad = \quad \left\{ \begin{array}{l} \langle(\mathsf{Bel}\ A\ \varphi), (\mathsf{M\text{-}Bel}\ \Theta\ \varphi)\rangle, \\ \langle(\mathsf{Bel}\ A\ \mathsf{AG}\neg\varphi), (\mathsf{M\text{-}Bel}\ \Theta\ \mathsf{AG}\neg\varphi)\rangle \end{array} \right\}$$

Minimal Social Commitment

$$(\text{J-Commit}\ \Theta\ \varphi\ \psi_{soc}\ \lambda_{soc}) \equiv$$

$$\forall A,\ A \in \Theta : [\neg(\text{Bel}\ A\ \varphi) \wedge (\text{Bel}\ A\ \text{EF}\varphi)] \wedge$$

$$\text{A}\left[\left(\begin{array}{l} (\text{Int}\ A\ \varphi)\wedge \\ ((\text{Bel}\ A\ \varphi) \Rightarrow \text{A}((\text{Int}\ A(\text{M-Bel}\ \Theta\ \varphi))) \curvearrowleft \chi \wedge \\ ((\text{Bel}\ A\ \text{AG}\neg\varphi) \Rightarrow \text{A}((\text{Int}\ A(\text{M-Bel}\ \Theta\ \text{AG}\neg\varphi))) \curvearrowleft \chi \end{array}\right) \curvearrowleft \chi\right]$$

where $\chi \equiv ((\text{M-Bel}\ \Theta\ \varphi) \vee (\text{M-Bel}\ \Theta\ \text{AG}\neg\varphi)))$

Mutual Belief ?

Definition 1:

$$(\text{M-Bel}\ \Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{Bel}\ A\ (\text{M-Bel}\ \Theta\ \varphi))$$

Definition 2:

$$(\text{E-Bel}^{0}\Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{Bel}\ A\ \varphi)$$

$$(\text{E-Bel}^{k}\Theta\ \varphi) \equiv \forall A,\ A \in \Theta\ (\text{E-Bel}^{k-1}\ \Theta\ \varphi)$$

$$(\text{M-Bel}\ \Theta\ \varphi) \equiv \forall m \in N(\text{E-Bel}^{m}\Theta\ \varphi)$$