Modèles d'agents : aspects théoriques liés aux intentions et aux interactions

Bernard ESPINASSE
Aix-Marseille Université (AMU)
LSIS UMR CNRS 7296
2012

- **Introduction**
- **Les agents : Systèmes intentionnels**
- **Théories de l'agent liées aux INTENTIONS**
- **Théories sociales liées aux INTERACTIONS**

Plan

Références bibliographiques

Cours :

- **Gleize M.P.**, Cours "Intelligence collective", Université de Toulouse, IRIT.
- **Quinqueton J.**, "Systèmes multi-agents", Université de Montpellier, LIRMM.
- **Esfandiari B.**, "Sofware Agents" Course, University of Carleton, Canada.
- **Florea A.M.,** "Agents et Systèmes Multi-agents", Université de Bucarest, Roumanie.

Articles :

- **Cohen, P. R. and Levesque, H. J.**, 1990a. "Intention is choice with comitment". Artificial Intelligence, 42, pp. 213-61.
- **Cohen, P. R. and Levesque, H. J.**, 1990b. "Rational interaction as the basis for communications". Intentions in Communication, pp. 221-56, MIT Press.
- **Rao, A. S. and Georgeff, M. P.**, 1991. "Modelling rational agents within a BDI architecture". In Proc. of Knowledge Representation and Reasoning (KR&R-91), pp. 473-484, Morgan Kaufman.
- **Rao, A. S. and Georgeff, M. P.**, 1995. "BDI Agents: From Theory to Practice". In Proc. of ICMAS-95, San Fransisco, June 1995.
- **Shoham, Y.**, 1993. "Agent Oriented Programming", Artificial Intelligence, 60(1), pp. 51-92.
- **Woolridge, M. and Jenning, N. R.**, 1995. "Intelligent agents: theory and practice". The Knowledge Engineering Review, 10(2), pp. 115-52.

Références bibliographiques

Livres :

- **Ferber J. (95), Les systèmes multi-agents**, InterEditions.
- **Weiss G. - editor (00), Multiagent Systems**, MIT Press.
- **Singh M. (94), Multiagent Systems,** Springer Verlag.
- **Conte R., Castelfranchi C. (1995), Cognitive and Social Action**, UCL Press.
- **Haddadi A. (95), Communication and Coopération in Agent Systems**, Springer Verlag.
- **Dennett, D. C.**, 1987. "The intentional stance", MIT Press.
- **O'Hare G.M.P. & Jennings N.R. - editors (96), Foundations of Distributed Artificial Intelligence**, Wiley-Interscience.
- **Bradsham M. - editor (97), Software Agents**, AAAI Press - The MIT Press.
- **Huhns M.N. & Singh M.P. - editors (97), Readings in Agents**, Morgan-Kaufmann.
- ...

1. Introduction

- **Propriétés des agents**
- **Les agents et systèmes intentionnels**
- **Anthropomorphisme ?**
- **Différentes théories des agents : liées aux intentions, aux interactions**
- **Principales contributions théoriques**

Propriétés des agents

Propriétés primaires des agents (agent = système informatique (logiciel)):

- **Autonomie :** opère sans intervention directe d'être humain ou autre et a un certain contrôle sur ses actions et ses états internes
- **Comportement social :** interagit avec d'autres agents (éventuellement humains) via un langage de communication agent (ACL)
- **Réactivité :** perçoit son environnement (monde physique, utilisateur via une interface graphique, collection d'autres agents, Internet ou tous à la fois) et répond aux changements qui apparaissent
- **Comportement intentionnel (pro-activeness):** n'agit pas seulement en réponse à son environnement, peut avoir un comportement dirigé vers un **but** et prendre des **initiatives**

Propriétés complémentaires des agents :

- **mobilité** (capacité): se déplacer dans un réseau informatique
- **véracité** (conjecture): selon laquelle un agent ne communique pas de mauvaises informations sans le savoir
- **bénévolat** (conjecture): les agents n'ont pas de buts incompatibles, et que chaque agent essaiera de faire ce qu'on attend de lui
- **rationalité** (conjecture): un agent agira de sorte à atteindre ses objectifs, au moins dans la limite de ses convictions

Les agents et systèmes intentionnels

- **psychologie populaire** : comportement humain caractérisé par des **attitudes mentales** (convictions, désirs, espoirs, craintes,...)
- "**Système Intentionnel**" [Dennett 87]: entité "dont le comportement peut être prédit en attribuant des convictions, des désirs, de la perspicacité rationnelle":
  - **de 1° ordre** : possède des convictions et des désirs (etc...) mais pas de convictions ni de désirs sur ses convictions et ses désirs...
  - **de 2° ordre** : possède des convictions et des désirs sur ses convictions et ses désirs, tant envers les autres qu'envers lui-même

Agent = système intentionnel

décrit par les **attitudes mentales** ou **intentionnelles** [Shoham & Cousins 94]:

- **informationnelles (information attitudes)** liées au monde dans lequel l'agent réside:
  => connaissances** et **croyances
- **de motivation (pro-attitudes)** guidant les actions de l'agent :
  => désirs, intentions, obligations, engagements
- **sociales** : ...
- **émotionnelles** : ...

Anthropomorphisme ?

pour **McCarthy** il existe des cas pour lesquelles **le point de vue intentionnel** est approprié pour des **systèmes artificiels** :

"**Attribuer des convictions, un libre arbitre, des intentions, une conscience, des compétences ou des désirs à une machine est légitime lorsqu'une** telle attribution exprime la même information pour la machine que pour une personne. Cela est utile lorsque l'attribution aide à comprendre la structure de la machine, son comportement passé ou futur, ou comment le réparer ou l'améliorer.

Ce n'est peut-être jamais logiquement nécessaire même pour des humains, **mais il se peut qu'exprimer de façon raisonnablement brève ce qui est réellement connu sur l'état de la machine dans une situation particulière nécessite des qualités mentales ou des qualités isomorphes à ces dernières**.

Les théories du croire, du connaître et du vouloir peuvent être construites pour des machines selon des paramètres plus simples que pour les humains, et ensuite applicables aux humains.

L'attribution de qualités mentales est plus directe pour des machines dont on connaît la structure comme les thermostats ou les systèmes d'exploitation informatiques, mais est plus utile quand elle s'applique à des entités dont la structure n'est pas totalement connue."

Théories des agents

Agents = systèmes intentionnels

Aspects individuels:

- Comment spécifier un agent capable d'agir de façon autonome ?
- Quelles propriétés notamment intentionnelles doit avoir un agent ?
- Comment représenter de manière formelle et raisonner sur ces propriétés ?, ...

Aspects collectifs:

- Comment faire coopérer une société d'agents
- Organisation, communication, coordination, ...

Théories des agents :

- **Représentation de notions intentionnelles ?**

⇒ *proposer des **formalismes** pour spécifier les agents, représenter leurs propriétés*

- **Raisonnement sur les notions intentionnelles ?**

⇒ *développer des **théories** utilisant ces formalismes et généralisant ces propriétés*

Principales contributions théoriques

Théories de l'agent liées aux INTENTIONS

- Moore 90 : Connaissance et action
- Méta-langages et modalités syntaxiques
- Singh 90 : logiques pour représenter intentions, croyances, connaissance, savoir-faire et communication
- Werner 88 : Modèle d'agent
- **Hintikka 62 : Les mondes-possibles - problème de l'omniscience**
- **Cohen Levesque 90 : pro-attitudes - buts et désirs**
- **Rao et Georgeff 91 : BDI architecture - croyances, désirs et intentions**
- **Shoham 93: AOP (Agent Oriented Programming) - croyances, compétences, engagements**
- Wooldridge 92 : logiques pour représenter les propriétés des SMA, …

Théories sociales des SMA liées aux INTERACTIONS

- **actes du langage** : Austin - Searle - Vandervecken - Cohen - Perrault - Levesque
- **maximes de la conversation** : Grice
- **dépendances - pouvoir** : Castelfranchi 90, Sichman 95
- théorie des jeux : Genesereth, Ginsberg, Rosenschein 86
- ...

2. Théories de l'agent liées aux intentions

- **Représentation des intentions**
- **Représentation des intentions : logiques des croyances**
- **Les mondes possibles [Hintikka 62]**
- **AOP - Agent 0 [Shoham 90]**
- **Théorie de l'intention [Cohen & Levesque 90]**
- **Architecture BDI [Kinny & Rao 91]**

Intentionnalité & intention

Intentionnalité : qu'est-ce qui fait qu'un agent fait ce qu'il fait?

Intention:

- Thomas d'Aquin : "**acte de volonté précédé d'un acte par lequel la raison ordonne quelque chose à sa fin**"
- Philosophie du moyen âge : "**la conscience s'appliquant à quelque chose d'autre qu'elle même**"

Cogniton **(atome cognitif [Ferber]) : unité cognitive élémentaire, structure calculatoire à propos de quelque chose, douée d'intentionnalité**

État mental : ensemble des cognitons

Types de cognitons :

- **interactionnels :** liés à **l'interaction** avec l'environnement, sa **perception**
- **représentationnels :** liés à la **représentation** du monde qui est construite
- **conatifs :** liés à **l'impulsion** déterminant l'action, **l'effet** produit sur un destinataire
- **organisationnels :** liés à la façon de **s'organiser** dans le raisonnement, l'action

Types des cognitons [Ferber]

Cognitons interactionnels	Cognitons conatifs
• **percepts** • **information** (croyance véhiculée par message) • **commande** • **requête** • **normes**	• **tendance/ but**: résultats de pulsions et demandes • **pulsion**: source interne de buts • **demande**: source externe de buts • **intention**: ce qui meut un agent • **engagement**: dépendance
Cognitons représentationnels	**Cognitons organisationnels**
• **croyance**: état du monde du point de vue d'un agent • **hypothèse**: représentations possibles du monde et des autres agents	• **méthode** : techniques, règles et plans pour atteindre un but (le "comment") • **tâche** : ce qu'il faut faire (conceptuel), ou bien déroulement d'une méthode

Types des cognitons [Ferber]

Système interactionnel (basé sur la perception):

- **perception passive** : traitement de signal, capteurs,...
- **perception active** : fait intervenir la représentation et les buts

Système représentationnel :

- **connaissance**: croyance, savoir
- **symbole / interprétation**: les mondes possibles
- **formaliser les croyances**: non monotone
- **savoir**: croyance vraie

$$sait(j, \alpha) : croit(j, \alpha) \wedge vrai(\alpha)$$

Que croire?

- croyances **environnementales** (perception, ...)
- croyances **sociales** (sur les groupe)
- croyances **relationnelles** (sur les autres)
- croyances **personnelles** (sur lui même)

Représentations des notions intentionnelles

soit l'assertion :

"Janine croit que Cronos est le père de Zeus"

traduction en logique classique du 1° ordre :

```
croît(Janine, Père(Zeus, Cronos))(1)
Zeus = Jupiter(2)
```

de (1) et (2) on en déduit :

"Janine croit que Cronos est le père de Jupiter", soit :

```
croît(Janine, Père(Jupiter, Cronos))(3)
```

Limite de la logique du 1° ordre « problème syntaxique » :

le 2° argument du prédicat `croît` est une formule de logique du 1° ordre et non un terme

=> (2) n'est pas une formule bien écrite dans la logique classique du 1° ordre

Limites de la logique classique (2)

Limite de la logique du 1° ordre « problème sémantique » (opacité inférentielle) :

```
croît(Janine, Père(Jupiter, Cronos))(3)
```

- les constantes `Zeus` et `Jupiter` se réfèrent exactement au même individu : le dieu suprême du monde antique.

Mais intuitivement :

"croire que le père de `Zeus` est `Cronos`"

≠

"croire que le père de `Jupiter` est `Cronos`"

=> on ne peut accepter (3)

Limites de la logique classique (2)

Pourquoi ?

- **les notions intentionnelles (convictions, désirs) sont relativement opaques** :
  - elles établissent des contextes opaques dans lesquels les règles de substitutions de la logique du 1°ordre ne s'appliquent pas
- **en logique classique la valeur sémantique d'une expression** dépend uniquement des représentations de ses sous-expressions :
  - Ex: la représentation de $p \wedge q$ est une fonction des valeurs exactes de p et de q (opérateurs fonctionnels exacts)
- **au contraire les notions intentionnelles comme les convictions ne sont pas fonctionnelles exactes**:
  - la valeur exacte de **"Janine Croît p"** ne dépendant pas uniquement de la valeur exacte de **p**

=> recherche d'autres logiques

Dépassement des limites de la logique classique

Dépassement du problème syntaxique :

- **langage modal** contenant des opérateurs modaux non fonctionnels exacts qui s'appliquent à des formules
- **méta-langage** : langage de 1° ordre à niveaux multiples contenant des termes qui représentent des formules d'autres langages. Notions intentionnnelles représentées par des prédicats du méta-langage avec tous les axiomes nécessaires

Dépassement du problème sémantique :

- **mondes possibles** (Hintikka 1962) : convictions, connaissance, buts... d'un agent = ensemble de mondes possibles supervisé par une relation d'accessibilité:
  - **avantages** : outils mathématiques liés à des théories de correspondance associées à la sémantique de ces mondes possibles
  - **inconvénients** : nombreuses difficultés dont le problème de l'omniscience logique, qui implique que les agents raisonnent parfaitement
- **structures symboliques interprétées ou phrasales:** convictions considérées comme des formules symboliques explicitement représentées dans une structure de données associées à un agent (Belief Data Structures) : un agent croit s'il se trouve dans sa structure de données de convictions

Pour une véritable théorie de l'agent

- **doit concerner toutes les attitudes mentales de l'agent**

- **devra définir comment les attributs de l'agent sont reliés et évolus :**
  - comment une **information** sur un agent et des **pro-attitudes** sont **reliées**?
  - comment un **état mental** d'un agent **change** au cours du temps ?
  - comment **l'environnement affecte** un **état mental** d'un agent ?
  - comment une **information** sur un agent et des **pro-attitudes** le conduise à rendre ses **actions plus performantes** ?

Contributions à une théorie de l'agent

- **Logiques des croyances** : convictions, connaissance, buts... d'un agent = ensemble de mondes possibles supervisés par une **relation d'accessibilité**
- **Hintikka, Kirple - Les Mondes Possibles** : convictions, connaissance, buts... d'un agent = ensemble de mondes possibles supervisé par une **relation d'accessibilité**
- **Cohen et Levesque - Théorie de l'intention** :
  - formalisme et théorie de l'intention basée sur **2 attitudes : convictions et buts**
  - utile pour raisonner sur les agents et utilisée dans l'analyse de conflit, la coopération entre plusieurs agents et la résolution coopérative de problèmes
- **Shoham** : **AOP** (Agent Oriented Programming)
- **Rao et Georgeff - Architecture BDI (Belief, Desire, Intention) :**
  - cadre logique pour la théorie de l'agent basé sur **3 attitudes: convictions, désirs et intentions**
  - formalisme est basé sur un modèle dans lequel les mondes accessibles des convictions, du désirs et de l'intention se ramifient en des structures temporelles
  - tente de répondre à : comment les convictions d'un agent quant au futur affectent ses désirs et ses intentions ?
  - considère **l'intérêt des plans sociaux** dans leur formalisme

Autres contributions à une théorie de l'agent (1)

Moore - Connaissance et action :

logique pour spécifier l'agent notamment les préconditions de connaissance pour ses actions : "de quoi un agent a-t-il besoin pour être capable de rendre certaines actions plus performantes? "

- **modèle de la compétence dans une logique** contenant une modalité pour la connaissance et un mécanisme logique dynamique pour modéliser des actions
- **agent peut utiliser de l'information incomplète** sur les moyens d'atteindre son but et rendre certaines actions plus performantes pour atteindre ce but

Singh :

logique pour représenter les intentions, les convictions, les connaissances, le savoir-faire et la communication dans un cadre ramificateur temporel (formalisme riche mais complexe permettant d'établir des propriétés)

Contributions à une théorie de l'agent (2)

Werner:

fondations d'un modèle général d'agent dans des domaines aussi variés que l'économie, la théorie des jeux, la sémantique de situation et la philosophie

Wooldridge - Modélisation des systèmes multi-agents :

- **objectif** : pas développer un cadre général pour une théorie de l'agent **mais élaborer des formalismes pouvant être utilisés dans la spécification et la vérification de SMA réels**
- **logiques pour la représentation des propriétés des SMA** permettant un modèle simple et général des SMA
- a montré comment les historiques obtenus au cours de l'exécution d'un SMA pourraient être utilisés comme le **fondement sémantique de logiques des convictions temporelles** tant linéaires que se ramifiant dans le temps

Représentation des intentions : logiques des croyances

Pour qu'un agent puisse raisonner :

- sur sa **propre connaissance**
- sur la **connaissance des autres agents**

=> **besoin de formalismes pour décrire ce qu'un agent "sait" et "croît" :**

D'où 2 opérateurs :

- **Belief** (croire) - B
- **Know** (savoir) - K

- **Pour exprimer que l'agent α croît une proposition P** (P = Paul est le père d'Alice)**:**

$$B(\alpha, Père(Paul, Alice))$$

- **Pour exprimer que l'agent α sait une proposition P et que P est supposée vraie :**

$$K(\alpha, Père(Paul, Alice))$$

Logiques des croyances : sémantique des atomes de croyances

la **valeur de vérité d'une croyance** (Belief) ne doit pas être la même quand un **terme équivalent** lui est **substitué** :

$$B(\alpha, Faim(Paul)) \neq B(\alpha, Faim(Père(Alice)))$$

même si Paul est le père d'Alice

pour définir la sémantique de Belief on étend le domaine à :

- un **ensemble d'agents**
- **chaque agent α**:
- un **ensemble de base de croyance $\Delta\alpha$** et
- un **ensemble de règles d'inférences $\rho\alpha$**
- une **théorie d'agent** comprend tout ce qui peut être dérivé à partir de ces règles

=> limitation du nombre et des types de déductions qui pourront être faites

Ainsi :

$B(\alpha,\phi)$ est vraie ssi ϕ est dans la théorie associée à l'agent α

Les mondes possibles [Hintikka 62]

Origines :

- au départ proposé par Hintikka en 1962
- actuellement formalisé en logique modale développée par Kripke en 1963

Exemple de monde possible : le jeu de poker

- la connaissance complète des mains des opposants est impossible à déterminer
- la **capacité à jouer** est **déterminée partiellement** par les **croyances** de l'agent sur les mains des opposants
- supposons que l'agent a un **As de Pique**, il faut :
  - **calculer toutes les possibilités** de distribution des cartes aux opposants : ce sont les **mondes possibles**
  - **éliminer les mondes qui ne sont pas possibles à partir de ce que sait l'agent**: ce qui reste = alternatives épistémiques (mondes possibles sachant certaines croyances). Quelque chose VRAI dans tous les mondes est dit CRU par l'agent (il est VRAI que l'agent a l'As de Pique)

Les mondes possibles [Hintikka 62]

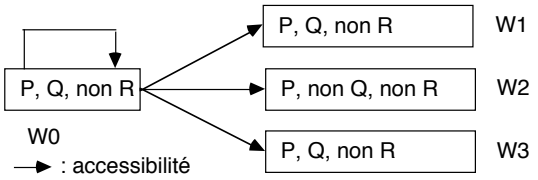
W0, W1, ..., Wn = des **mondes possibles**

langage similaire à la logique des croyances : logique classique du 1° ordre **+ opérateur modal K**

Sémantique pour les formules bien formées (fbf) : une fbf n'est pas vraie ou fausse, mais vraie ou fausse par rapport à un monde possible : il y a donc une interprétation par monde possible

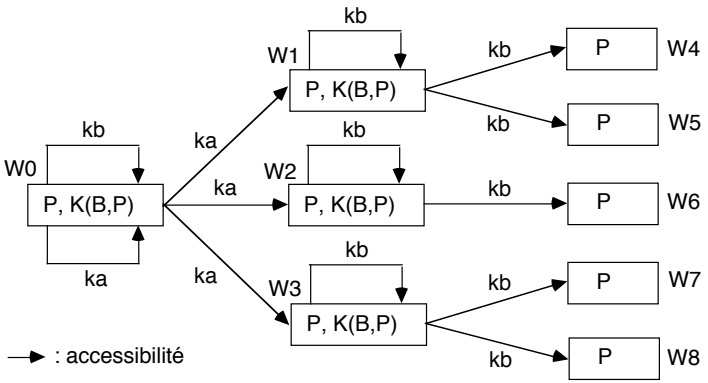
Relation d'accessibilité k :

- k(α, Wi, Wj) : le monde Wj est accessible à partir du monde Wi pour l'agent α
- K(α,ϕ) : ϕ a la valeur VRAIE dans le monde Wi ssi ϕ a la valeur VRAIE dans tous les mondes accessibles à partir de Wi pour l'agent α :



l'agent α sait P et non R mais ne sait pas Q

La logique des mondes possibles (Hintikka 1962)



Propriétés de K

- **AXIOME 1 : axiome de distribution :**

  $(K(\alpha,\phi) \wedge K(\alpha,\phi \Rightarrow \psi)) \Rightarrow K(\alpha,\psi)$ ou $K(\alpha,\phi \Rightarrow \psi) \Rightarrow (K(\alpha,\phi) \Rightarrow \psi)) \Rightarrow K(\alpha,\psi))$

- **AXIOME 2 : axiome de la connaissance :**

  $K(\alpha,\phi) \Rightarrow \phi$

- **AXIOME 3 : axiome de l'introspection positive :**

  $K(\alpha,\phi) \Rightarrow K(\alpha, K(\alpha,\phi))$

- **AXIOME 4 : axiome de l'introspection négative :**

  $\neg K(\alpha,\phi) \Rightarrow K(\alpha, \neg K(\alpha,\phi))$

- **Règle 1 : la nécessitation épistémique :**

  from $\vdash \phi$ infer $K(\alpha,\phi)$

- **Règle 2 : l'omniscience logique :**

  from $\phi \vdash \psi$ and from $K(\alpha,\phi)$ infer $K(\alpha,\psi)$

Propriété de B

- **AXIOME 8 :**

  $\neg B(\alpha, false)$

- **AXIOME 9 : axiome de l'introspection positive :**

  $B(\alpha,\phi) \Rightarrow B(\alpha, B(\alpha,\phi))$

et on peut dire que :

$B(\alpha,\phi) \Rightarrow \mathbf{K}(\alpha, B(\alpha,\phi))$

$B(\alpha, B(\alpha,\phi)) \Rightarrow B(\alpha,\phi))$

$B(\alpha 1, B(\alpha 2,\phi)) \Rightarrow B(\alpha 1,\phi))$

Omniscience de la logique

A partir de :

$\phi \vdash \psi$ et de $K(A,\phi)$ on infère : $K(A,\psi)$

d'où un agent connaît toutes les conséquences de sa connaissance

=> irréaliste de supposer cela pour des agents

AOP - Agent 0 [Shoham 93]

- La **programmation orientée agents** a été proposée par Shoham comme **un nouveau paradigme de programmation** (AOP - Agent Oriented Programming), une spécialisation du paradigme de programmation orientée objets
- **Agent0 =** un langage de programmation de ce nouveau paradigme
- Agent0 est basé sur modèle de **logique modale** définissant la sémantique des constructions du langage avec des **opérateurs modaux**
- Les agents AGENT0 agissent dans un environnement où le **temps est explicité, discret** et **linéaire**

Agent0 est un formalisme basé sur:

- **2 modalités de base** : la **croyance** (**B**) et l'**engagement** **(Commitment – CMT)**
- une **représentation explicite du temps** incorporée.

AOP - Agent 0 [Shoham 93]

Les agents Agent0 agissent dans un environnement où le **temps est discret** et **linéaire** et ils sont caractérisés par les notions mentales suivantes :

- **B** = **croyance** = prédicat à 2 arguments :

  $\mathbf{B}_x\ p^t$ **:** au temps t, l'agent x croit la proposition p (définie récursivement)

  Ex : $(B_{Marie}\ (B_{Georges}\ aime(Marie, Denis)^2)^7)^{10}$

  signifie qu'au temps 10, Marie croit que Georges croit au temps 7, que Marie aime Denis au temps 2.

- **CMT** = **engagement** (commitment)= prédicat à 3 arguments:

  $\mathbf{CMT}_{x,y}\ j^t$ **:** l'agent **x** est engagé face à l'agent **y** pour accomplir l'action **j**.

  Ex : $(CMT_{Georges, Marie}\ aime(Marie, Georges)^{10})^2$

  signifie qu'au temps 2, Marie est engagée face à Georges de l'aimer au temps 10.

- **DEC = décision :** $\mathbf{DEC}_x\ j^t =_{def} \mathbf{CMT}_{x,x}\ j^t$ = à l'instant t l'agent x a décidé de faire l'action j (engagement à l'instant t de agent x envers lui même de faire j)
- **CAN** = **habilités :** $\mathbf{CAN}_x\ j^t$ = à l'instant t, l'agent x est capable de faire l'action j.

Langage Agent0

Les composantes principales du langage Agent0 sont :

- **instructions factuelles** : `(t (aime alice paul))` : il est vrai qu'Alice aime Paul
- **instructions pour l'exécution des actions privées :** `(DO t1 fermer)` : l'agent exécute l'action privée « fermer » à l'instant t1 ;
- **instructions pour des actions communicatives** :
  - `(INFORM t1 b fait)` : l'agent informe un autre agent b qu'il croit que le fait est vrai à l'instant t1;
  - `(REQUEST t1 b action)` : l'agent demande à l'agent b d'exécuter à l'instant t1 ;
  - `(UNREQUEST t1 b action)` : l'agent retire sa demande à l'agent b d'exécuter l'action à l'instant t1 ;
  - `(REFRAIN action)` - l'agent ne doit pas s'obliger à exécuter l'action ;
- **représentation des conditions mentales dans le langage :**
  - `(B (t fait))` **:** représente $\mathbf{B}_x\ a^t$ où x=fait; l'agent x est implicite dans cette représentation car il s'agit de son programme ;
  - `((CMT t b) action)` : représente $\mathbf{CMT}_{x,y}\ j^t$, où j=action ;
  - `((DEC t) action)` : représente $\mathbf{DEC}_x\ j^t$, où j=action ;
  - `(CAN t) action)` : représente $\mathbf{CAN}_x\ j^t$, où j=action ;

Langage Agent0

instructions pour des actions conditionnées : la forme générale de ces instructions est (IF condition-mentale action), par exemple :

```
(IF (B (t (valeur actions 100)))
        (INFORM t1 b (t (valeur actions 100))))
```

règles d'engagement; indiquent les conditions sous lesquelles un agent assume un certain engagement ; la forme générale d'une règle d'obligation est :

```
(COMMIT condition-message condition-mentale (agent action))
```

- *« condition mentale » cf plus haut*
- *« condition du message » a pour rôle d'identifier le type de message reçu et l'action spécifiée dans le message,*
- *« agent » est l'agent envers lequel on prend l'engagement.*

Une règle d'engagement dit **qu'un agent s'engage à faire une certaine action** si :

- la condition message est satisfaite par le message reçu ;
- la condition mentale est vraie dans l'état courant de l'agent ;
- l'agent est capable de faire l'action spécifiée dans la règle ;
- l'agent n'a pas l'interdiction d'exécuter l'action (REFRAIN)

Langage Agent0

Exemple de règle d'engagement est :

```
(COMMIT (?b REQUEST ?action)
     (B (now (mon_ami ?b)))
     (?b ?action))
```

où par **?b** on indique un **agent quelconque** (le préfixe ? indique les variables du langage) ;

cette règle signifie :

SI

- l'agent **reçoit un message de demande** (`REQUEST`) pour **effectuer une certaine action** (`?action`), et
- **s'il croit qu'à l'instant présent** (`now`) que `l'agent ?b` **est son ami**

ALORS

- notre **agent s'oblige envers** `?b` **à exécuter l'action** `?action`.

Exemple de message de demande : `(Alice REQUEST achete_livre),` notre agent va s'obliger envers Alice (?b), si Alice est son amie, d'exécuter pour lui l'action achete_livre (?action).

Langage Agent0

Comparaison programmation orientée objets / programmation orientée agents [Shoham 93] :

	POO	POA
Unité de base	Objet	Agent
Paramètres définissant l'état de l'unité de base	Pas de contraintes	Croyance, décisions, obligations, habilités
Processus de calcul	Envoi de message, méthodes pour la réponse	Envoi de message, méthodes pour la réponse
Types de messages	Pas de contrainte	Informer, demander, offrir, promettre, accepter, rejeter
Contraintes sur les méthodes	Pas de contraintes	Consistance, vérité

Théorie de l'intention [Cohen & Levesque 90]

Bratman, Cohen et Levesque identifient 7 propriétés que doit satisfaire une théorie raisonnable de l'intention :

1. les **intentions posent des problèmes** aux agents, lesquels ont besoin de déterminer des **moyens** de les **concrétiser**
2. les **intentions** sont un "**filtre**" pour **adopter d'autres intentions** qui ne doivent pas rentrer en conflit
3. les **agents suivent les succès** de leurs intentions et ils les réessayer en cas d'échec
4. les **agents croient** que leurs **intentions** sont **possibles**
5. les **agents ne croient pas** qu'ils vont **provoquer** leurs **intentions**
6. dans certaines circonstances, les **agents croient** qu'ils vont **provoquer** leurs **intentions**
7. les **agents** n'ont **pas besoin d'envisager** tous les **effets attendus** de leurs **intentions**

Théorie de l'intention [Cohen & Levesque 90]

Définition de l'intention retenue :

on dit qu'**un agent x a l'intention de faire une action a**, noté **`intention(x, a)`**, si :

- **x** a comme **but** qu'une **proposition p** portant sur un état du monde soit **vrai**, noté **`but(x, p)`**, et
- que les **conditions suivantes** sont **vérifiées** :
  - **x** croit que **p** appartient aux **conséquences** de **a**
  - **x** croit que **p** n'est **pas vrai actuellement**
  - **x** croit qu'il est **capable** de faire **a**
  - **x** croit que **a** sera **possible** et donc que **p pourra être satisfait**

Théorie de l'action rationnelle :

- permet de lier les notions **d'intentions**, de **croyances** et les **actions**
- introduction de **structures abstraites** dérivées constituant une théorie partielle de l'action rationnelle ("partial theory of rational action"):
- fournit une **notation uniforme** et **pratique**
- repose sur la **logique modale** et **des mondes possibles**: logique 1°ordre + **opérateurs** de modalités

Théorie de l'intention : Notation [Cohen & Levesque 90]

Modalités de base (inspiré de la logique des mondes possibles):

- **BELief**: **croyance** (**Croire**)
- **GOAL**: **but - intention** (**But**)

Modalités pour la gestion d'actions (action = événement qui rend possible la satisfaction d'une proposition) :

- **HAPPENS**: **une action va se passer** (**VaSePasser**)
- **DONE**: **une action vient de se s'achever** (**Achevée**)

Sachant :

- x : un agent
- p : une proposition
- a, b : 2 actions, chacune est considérée comme une séquence d'évènements

On a :

- (**BEL x p**) **x** a **p** comme **croyance** (**Croire x p**)
- (**GOAL x p**) **x** a **p** comme **but** (**But x p**)
- (**HAPPENS a**) une action **a va se passer** (**VaSePasser a**)
- (**DONE a**) une action **a vient de s'achever** (**Achevée a**)

Syntaxe de la théorie de l'intention

Sachant :

- x : un agent
- p : une proposition
- a, b : 2 actions, chacune est considérée comme une séquence d'évènements
- e : une variable événement

Connecteurs de structuration des séquences d'évènements (inspiré de la logique dynamique d'Harel):

- (**AGT x e**) **x** est le seul **agent** de la **séquence d'événements e** (**Agent x e**)
- **e1 <= e2** **e1** est une **sous-séquence** initiale de **e2**
- **a;b** **composition séquentielle d'actions** (on fait a puis b)
- **a|b** **choix non déterministe** d'actions (on fait soit a soit b)
- **p?** est une **action qui teste si p est vrai**
- **a*** **répétition** : l'action se répète indéfiniment

Sémantique de la théorie de l'intention

- **monde possible** = séquence d'événements vers le passé ou vers le futur
- **la vérité d'une proposition dépend :**
  - du **monde** dans lequel elle est
  - de **l'instant dans le cours des événements** (repéré par un entier : index du temps)

- **2 relations d'accessibilité :**
  - **B(φ, x, t, σ*)** ce que croit l'agent x dans le monde σ^* est compatible avec ce que croit l'agent x dans le monde φ à l'instant t
  - (BEL x p) est vraie quand p est vraie dans tous les mondes reliés par B
  - **G(φ, x, t, σ*)** les buts de l'agent x dans le monde σ^* sont compatibles avec les buts l'agent x dans le monde φ à l'instant t
  - (GOAL x p) est vraie quand p est vraie dans tous les mondes reliés par G

Structures de contrôle plus élaborées

- **Action condition** : si p alors a sinon b (on effectue a dès que p est vérifié et b si p est faux)
  [IF p THEN a ELSE b] $=_{def}$ p?;a I ¬p?; b
- **Boucle tant-que** : on répète a tant que p est vrai
  [WHILE p DO a] $=_{def}$ (p?;a)*; ¬p?
- **Finalement (eventually)**: opérateur indiquant qu'il existe une séquence d'action e après laquelle p sera vrai
  EVENTUALLY (p) $=_{def}$ ∃a (HAPPENS a;p?)
- **Toujours** : opérateur signifiant que p est toujours vrai (p jamais faux)
  p $=_{def}$ **¬EVENTUALLY** (¬p)
- **Plus tard** : opérateur signifiant que p n'est pas vrai maintenant mais il sera vrai plus tard
  (LATER p) $=_{def}$ ¬p ∧ ◊p
- **Avant** : opérateur signifiant que p vient avant q si quand q est vrai p a été vrai précédemment
  (BEFORE p q) $=_{def}$ ∀c (HAPPENS c;q?) ⊃ a (a≤ c) ∧ (HAPPENS a;p?)
- **De même sont définis :**
  (DONE x a) = (DONE a) ∧ (AGT x a)
  (HAPPENS x a) = (HAPPENS a) ∧ (AGT x a)
- **Conséquence déduite : EVENTUALLY (p) ∧ (BEFORE p q) ⊃ EVENTUALLY (q)**
  (si q est finalement vérifiée et que p est avant q, alors p sera nécessairement vérifié)

Structures de contrôle plus élaborées

but persistant:

but que l'on conserve tant que certaines conditions demeurent :

- **(P-G0AL x p)** $=_{def}$ (GOAL x (LATER p)) ∧
  (BEL, x ¬p) ∧
  [BEFORE ((BEL x p) ∧ (BEL x ¬p)) ∧ ¬(GOAL x (LATER p))]
- **But-P(x, p)** $=_{def}$ But(x, PlusTard(p)) ∧
  Croit(x,¬p) ∧
  [Avant(Croit(x,p) ∨ Croit(x, Tjrs(¬p)), ¬But(x, PlusTard(p))]

intention :

engagement "fanatique" d'accomplir l'action **a** dès ,que l'agent croit qu'elle peut l'être :

- **(INTEND1 x a)** $=_{def}$ (P-GOAL x [DONE x (BEL x (HAPPENS a)) ?;a])
- **Intention-Fan(x, a)** $=_{def}$ But-P(x, Achevee(x, Croit(x, VaSePasser(a))?; a))

Structures de contrôle plus élaborées

Autres modalités définies :

- **(KNOW x p)** $=_{def}$ p ∧ (BEL x p)
- **(COMPETENT x p)** $=_{def}$ (BEL x p) ⊃ (KNOW x p)

- **(A-GOAL x p)** $=_{def}$ (GOAL x (LATER p)) ∧ (BEL x ¬p)
- **(P-G0AL x p)** $=_{def}$ (GOAL x (LATER p)) ∧ (BEL, x ¬p) ∧ [BEFORE ((BEL x p) ∧ (BEL x ¬p)) ∧ ¬(GOAL x (LATER p))]

- **(INTEND1 x a)** $=_{def}$ (P-GOAL x [DONE x [BEL x (HAPPENS a))?;a])
- **(INTEND2 x p)** $=_{def}$ (P-GOAL x ∃e[DONE x [(BEL x ∃e'(HAPPENS x e';p?)) ∧ ¬(GOAL x ¬(HAPPENS x e;p?))]?; e; p?))

- **(P-R-GOAL x p q)** $=_{def}$ (GOAL x (LATER p)) ∧ (BEL x ¬p) ∧ (BEFORE [(BEL x p) v (BEL x ¬p) v (BEL x ¬q)] ¬(GOAL x (LATER p)))
- **(INTEND1 x a q)** $=_{def}$ (P-R-GOAL x [(DONE x (BEL x (HAPPENS x a))?;a)] q)
- **(INTEND2 x a q)** $=_{def}$ (P-R-GOAL x ∃e[(DONE x (BEL x ∃e'(HAPPENS x e';p?)) ∧ ¬(GOAL x ¬(HAPPENS x e;p?))] q)

Architecture BDI (Belief-Desire-Intention) [Kinny & Rao 91]

- **Hyp. 1**: **l'agent** est immergé dans un **environnement** :
  - dont il reçoit des **événements** véhiculant des **informations** sur l'état de l'environnement
  - dans lequel il agit par des **actions** qui **modifie** cet environnement
- **Hyp. 2**: **l'agent** a des **représentations** (objets, structures de données...) de :

  Belief = ***Croyance*** : Les croyances d'un agent sont les informations que l'agent possède sur l'environnement et sur d'autres agents qui existent dans le même environnement.

  Desire = ***Désir*** : Les désirs d'un agent représentent les états de l'environnement, et parfois de lui-même, que l'agent aimerait voir réalisés.

  -> buts : sous-ensemble de désirs consistants que l'agent choisit

  Intention = ***Intention*** : Les intentions d'un agent sont les désirs que l'agent a décidés d'accomplir ou les actions qu'il a décidé de faire pour accomplir ses désirs.

Architecture BDI (Belief-Desire-Intention) [Kinny & Rao 91]

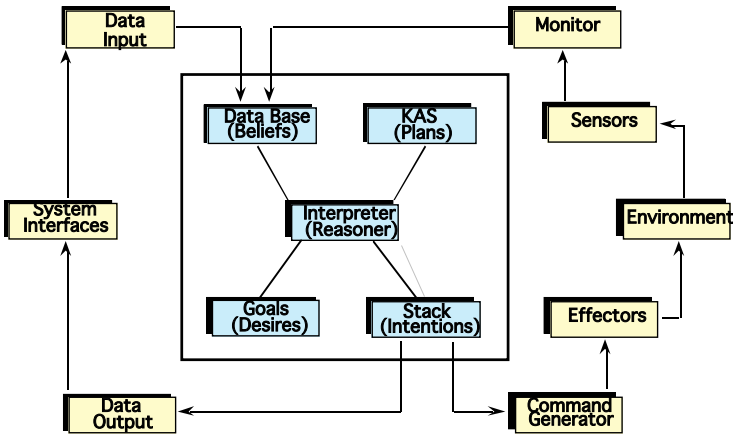
- **Hyp. 3**: l'agent a un **mécanisme de contrôle** qui s'assure que :
  - ses **croyances évoluent** en réponse aux **événements** externes (et parfois aux actions internes)
  - ses **intentions déterminent** et **engendrent** des séquences **d'action** à exécuter
  - ses **intentions évoluent** du fait de l'évolution de ses croyances, la satisfaction / l'échec de ses désirs, l'exécution d'action et la survenance de nouveaux événements

L'agent BDI

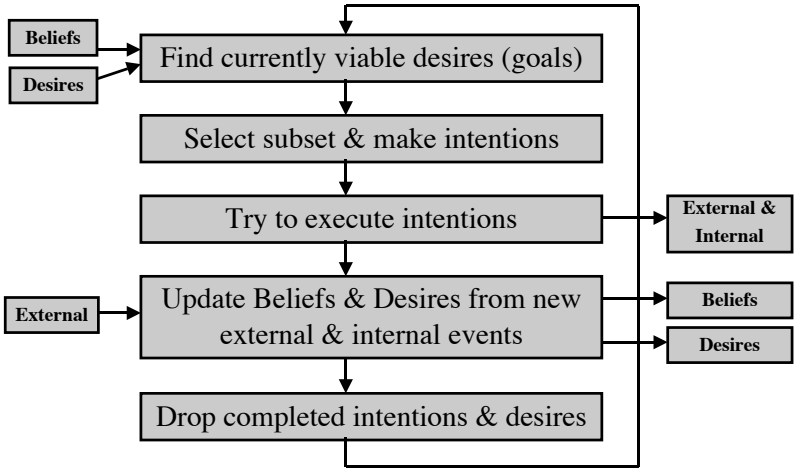
Beliefs	Desires
current state of the world (as seen by the agent) Variables, logical expressions, Database entries	goals, long term aims plans Priorities for plans Payoffs for plans

Intentions
instantiations of part(s) of plan fills in plans with sub-plans/actions alter beliefs & desires query tasks / resources delegate tasks to other agents

L'Architecture BDI : Architecture type des agents cognitifs

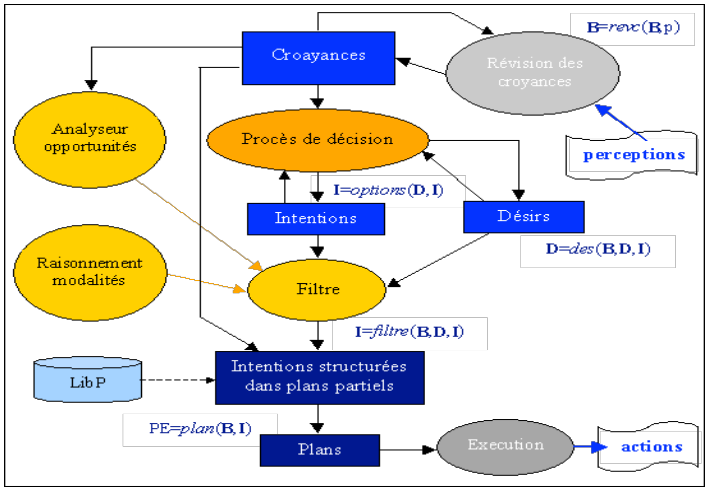


Exemple de boucle de fonctionnement d'un agent BDI



Exemple de boucle de fonctionnement d'un agent BDI

D'après Florea :



Exemple de boucle de fonctionnement d'un agent BDI

revc : **B** x P → **B** = fonction de **révision des croyances** de l'agent lorsqu'il reçoit de nouvelles perceptions sur l'environnement, où P = ensemble des perceptions de l'agent; elle est réalisée par la composante Révision des croyances;

option : **D** x **I** → **I** = fonction représentant le **processus de décision de l'agent** prenant en compte ses désirs et ses intentions courantes; cette fonction est réalisée par la composante Processus de décision;

des : **B** x **D** x **I** → **D** = fonction pouvant **changer les désirs d'un agent** si ses croyances ou intentions changent, pour maintenir la consistance des désirs de l'agent (on suppose dans notre modèle que l'agent a toujours des désirs consistants); cette fonction est également réalisée par la composante Processus de décision;

filtre : **B** x **D** x **I** → **I** = fonction la plus importante car **elle décide des intentions à poursuivre**; elle est réalisée par la composante Filtre.

plan : **B x I** → **PE** = fonction **transformant les plans partiels en plans exécutables**, **PE** étant l'ensemble de ces plans ; elle peut utiliser, par exemple, une bibliothèque de plans, représentée par le module **LibP** dans la figure. Un **plan** est une séquence d'actions à exécuter dans le temps

Algorithme de contrôle d'un agent BDI

Soient **B0**, **D0** et **I0** les croyances, désirs et intentions initiales de l'agent.

Algorithme de contrôle d'un agent BDI :

```
1 B←B0
2 D←D0
3 I←I0
4 répéter
   4.1 obtenir nouvelles perceptions p
   4.2 B←revc(B, p)
   4.3 I←options(D, I)
   4.4 D←des(B, D,I)
   4.5 I←filtre(B, D, I)
   4.6 PE←plan(B, I)
   4.7 exécuter(PE)
jusqu'à ce que l'agent soit arrêté
fin
```


BDI : Stratégies d'obligation

Obligation aveugle :

Un agent suivant cette stratégie va maintenir ses intentions jusqu'à ce qu'elles soient réalisées, plus précisément jusqu'à ce qu'il croie qu'elles sont réalisées.

Obligation limitée :

Cette stratégie dit que l'agent va maintenir ses intentions ou bien jusqu'à ce qu'elles soient réalisées ou bien jusqu'à ce qu'il croie qu'elles ne sont plus réalisables.

Obligation ouverte :

Un agent ayant une stratégie d'obligation ouverte maintient ses intentions tant que ces intentions sont aussi ses désirs;une fois que l'agent a conclu que ses intentions ne sont plus réalisables, il ne les considère plus parmi ses désirs.

3. Théories de l'agent liées aux interactions

- **Interactions intentionnelles et non intentionnelles**
- **Actes du langage [Austin, Searle, Vandervecken]**
- **Maximes de la conversation [Grice]**
- **Théorie de la dépendance [Castelfranchi 1990] et notation [Cohen & Levesque 90]**
- **Organisation, auto-organisation et émergence**

Interactions intentionnelles et non intentionnelles

- **interaction non intentionnelle :**

  => apparition dans le champ de perception de l'agent

- **interaction intentionnelle :**
  - **communication via l'environnement** : propagation systématique de stimuli => impossible de spéculer les destinataires : "**trace**" (phéromones)
  - **échanges, communication d'informations** à différents niveaux :
    - perception des actions
    - envoi de signaux
    - expression d'intentions, de croyances, de plans
  - **communication entre agents complexes** = résultats d'actions rationnelles, d'intentions : **dialogue**

=> ACTES DE LANGAGE (psychologie sociale)

Actes du langage [Austin, Searle, Vandervecken]

"Communiquer c'est agir" : la communication envisagée sous la forme d'une action, qui doit être gérée comme les autres actions

Catégoriser des types de communication:

- **informer,**
- **demande de faire,**
- **demande d'info,**
- **réponses,**
- **promesse,**
- **proposition et offre...**

Type de communication décrit par ses conséquences :

- permet un **traitement logique de la communication**
- **protocoles de communication** associés à chaque acte de langage

Actes du langage [Austin, Searle, Vandervecken]

Composantes d'un acte de langage (actes élémentaires):

- **locutoire**: génération des **énoncés** (production d'une phrase dans une langue donnée)
- **illocutoire**: **acte** réalisé par le **locuteur** sur le **destinataire** : **informer, demande de faire, demande d'info, réponses, promesse, proposition, offre...**

  Ex: affirmer("il pleut !") ou questionner("il pleut ?")

- **perlocutoire**: **effets** que les actes illocutoire peuvent avoir sur l'**état du destinataire**

  Ex: persuader = affirmer avec le désir que le destinataire partage les croyances du locuteur.

Actes du langage [Austin, Searle, Vandervecken]

Classification des actes du langage:

- **informatifs**: affirmer un fait - Ex : "Achille a 4 ans"
- **directifs** :
  - **exercitifs**: demander de faire quelque chose
  - **interrogatifs**: poser une question
- **promissifs**: engagement à accomplir un acte - Ex : je ferai cours demain
- **expressifs**: exprimer un état - Ex : je suis heureux
- **déclaratifs**: accomplir un acte par l'énonciation - Ex : je t'aime

Actes du langage : acte illocutoire

Acte illocutoire :

- produire un certain effet sur le **destinataire** lors de la formulation d'un **énoncé**
- largement étudiés en **pragmatique du langage**

Acte illocutoire = contenu propositionnel (P) + force illocutoire (F)

F(P) ou <performative> (<contenu>)

- **performative** = type d'acte illocutoire - verbe (Informer, Demander de faire, Questionner, Répondre, Promettre, Affirmer,...)

Exemples :

- **Affirmer (il pleut)**
- **Questionner (il pleut)**

Succès d'un acte illocutoire : acte reconnu par l'auditeur

Langage opératoire KQML

Maximes de la conversation [Grice]

Maximes de quantité : contribution soit aussi informative que nécessaire, mais pas plus

Maximes de qualité :

- ne pas dire ce que l'on croit être faux,
- ni ce que l'on n'a pas de raisons de considérer comme vrai

Maxime de relation : être pertinent

Maximes de manière :

- éviter l'ambiguïté
- être bref et ordonné

Pourquoi ?

- **objectifs**: **pertinence / efficacité / flexibilité**
- **non réduit** à une technique informatique
- **traitement formel de la communication en tant qu'acte**

Mais relativiser !

adapter aux capacités de raisonnement des agents

Théorie de la dépendance: adoption de buts, dépendance et pouvoirs sociaux [Castelfranchi 1990]

Pourquoi adopte-t-on les buts des autres ?

coopération, échange social : complémentarité des agents

- **Coopération :**

  Ex : i et **j** ont le même but **g** et ne peuvent pas l'atteindre seul

- **Échange social :**

  Ex :
  - **i a comme but g et j comme but g' et i peut aider j à atteindre g'**
  - **i peut proposer d'aider j pour atteindre g'**
  - **si en contrepartie j l'aide à atteindre g**
  - **et j peut accepter**

Notion de dépendance : agent **i** dépend d'un agent **j** pour atteindre un de ses propres buts

Notion de pouvoir : pouvoir d'influencer l'autre

Usage de la notation de Cohen & Levesque

Buts communs et parallèles : notations [Cohen&Levesque 90]

Sachant :

- x et y : 2 agents différents
- a : action;
- r : ressource;
- p : proposition (formule bien formée - fbf)

  - (**RESOURCE r a**) **r** est **nécessaire** pour **a**
  - (**CANDO x a**) **x** peut exécuter **seul a**
  - (**DONE_BY x a**) = def (DONE a) ∧ (AGT x a)
  - (GOAL x p) **x** a comme **but p**
  - (BEL x p) **x** croit **p**
  - (DONE a) l'action **a** est **achevée**
  - (HAPPENS a) l'action **a** va **être effectuée**

- **But commun :** (**I_GOAL X Y P**) $=_{def}$ (GOAL X P) ∧ (GOAL Y P)
- **But parallèle :** (**P_GOAL X Y P**) $=_{def}$ (GOAL X P*) ∧ (GOAL Y P*)

* qualifie P de **but parallèle** dans le sens que lorsque on analyse la proposition P on trouve que X et Y ont des rôles identiques

Adoption de buts : notations [Cohen&Levesque 90]

Résultat d'une stratégie d'influence

Coopération = i doit partager un but avec j par l'utilisation de stratégies d'influence

(**OBTAIN** x p) $=_{def}$ (GOAL x p) ∧ (**EVENTUALLY** (p ∧ (BEL x p)))

ADOPT(x y p) $=_{def}$ (GOAL x (**EVENTUALLY** (**OBTAIN** y p)))

d'où : (BEL x ((OBTAIN y p) > (OBTAIN x q))) => **(ADOPT x y p)**

Contradiction entre agent bienveillant / rationnel :

- **agent rationnel** : il associe un coût aux actions et considère que les ressources sont limitées
- un agent **non rationnel** : il entreprend des actions ou utilise des ressources pour permettre aux autres agents d'atteindre leurs buts

Types d'adoption de but

Adoption terminale: adoption de but de haut niveau

- **adoption personnelle** : adopter toujours les buts d'un agent donné adoption indépendante de la nature du but justifiée que par des relations amitié amour

  la bienveillance = adoption personnelle

- **adoption non-personnelle** : un agent adopte certains buts d'un ensemble d'agents bien déterminé, la nature du but est prise en compte :

  - **adoption fonctionnelle** : adoption de certains buts à cause de la **structure fonctionnelle** dans laquelle il est inséré
  - **adoption normative** : il n'existe pas de structure fonctionnelle ni d'adoption personnelle: adoption due à des **normes dans la société**

Adoption instrumentale : adoption du but d'un autre agent car il profite de ce fait

Dépendances & pouvoirs sociaux: notations [Cohen&Levesque 90]

Notions de pouvoir :

- **pouvoir de** : un agent **i** a le pouvoir de **g** (but) s'il peut atteindre **g**
- **pouvoir sur** : un agent **i** a le pouvoir sur un autre agent **j** (en ce qui concerne **g**) s'il peut **l'aider** ou **l'empêcher d'atteindre g**

Notion de dépendance :

Agent **i** est **dépendant** d'un agent **j** (en ce qui concerne **g**) :

- s'il n'a pas le pouvoir de **g** et **j** a ce pouvoir ou
- s'il a le pouvoir de **g** sauf si **j l'empêche** d'atteindre **g**

Dépendances & pouvoirs sociaux: notations [Cohen&Levesque 90]

Dépendances de ressources : R_DEP

- un objet ou un événement dans le monde externe augmente la probabilité qu'un certain état du monde soit atteint
- cet état du monde est représenté comme but d'au moins un agent

(**R_DEP x r a p**) $=_{def}$ (GOAL x p) ∧ (RESOURCE r a) ∧
((DONE-BY x a) ⊃ (EVENTUALLY p))

Dépendances sociales : S_DEP

- un agent peut entreprendre une action dont le résultat augmente la probabilité qu'un certain état du monde soit atteint
- cet état du monde est représenté comme but d'un autre agent qui pour sa part est incapable d'entreprendre l'action en question :

(**S_DEP x y a p**) $=_{def}$ (GOAL x p) ∧ ¬(CANDO x a) ∧ (CANDO y a) ∧
((DONE_BY y a) ⊃ (EVENTUALLY p))

Types de relation de dépendances

OU-dépendance :

- **alternatives de partenaires** : une action **a** nécessaire pour atteindre le but **g** peut être réalisée par plusieurs agents : Il suffit que l'un d'entre eux la réalise pour que **g** soit atteint
- **alternatives d'actions** : un but **g** peut être atteint en entreprenant des actions différentes **a, a', a"** réalisées par des agents différents : Il suffit que l'une de ces actions soit entreprise et **g** est atteint

ET-dépendance :

- **dépendance multi-partite** : un **ensemble d'actions est nécessaire** pour atteindre **g**
- **dépendance multi-but** : un agent **i** dépend de **j** pour atteindre plusieurs buts soit pour une action soit pour des actions différentes

la OU-dépendance diminue le degré de dépendance, la ET-dépendance l'augmente

Dépendance unilatérale :

(**DEP x y a p**) $=_{def}$ (GOAL x p) ∧ (¬(CANDO x a)) ∧
(CANDO y a) ∧ ((DONE a) > (OBTAIN x p))

Types de relation de dépendances (suite)

Dépendance mutuelle = coopération

- la **coopération** ne se **réduit pas à l'adoption de but,** pour qu'il y ait coopération il faut qu'il y ait **dépendance mutuelle entre les buts**
- **i** et **j** dépendent l'un de l'autre pour atteindre un même but **g** dont un plan pour le réaliser nécessite au moins 2 actions a_1 et a_2 : Chacun d'entre eux est capable de réaliser une seule de ces actions :

(**M_DEP x y p**) $=_{def}$ ∃a_x ∃a_y (DEP x y a_y p) ∧ (DEP y x a_x p) ∧
(S_DEP x y a_1 p) ∧ (S_DEP y x a_2 p)

Dépendance réciproque = échange social

i dépend de **j** pour un but **g** et **j** dépend de **i** pour un but **g'** et gl= g'

(**S_DEP x y a_1 p_1**) ∧ (**S_DEP y x a_2 p_2**)

De la dépendance de ressource à la dépendance sociale

SI **x** dépend d'une ressource **r** pour atteindre **p** et si **y** contrôle **r** ALORS **x** dépend de **y** pour **r**

Dépendance via influence

- power of influencing (**INFL_POWER x y a p**) : **x** peut influencer **y** s'il peut réaliser une action **a** qui fait que **y** a comme but **p**

Coopération

Formalisation de la coopération :

(**COOP_I x y p**) $=_{def}$ (BEL x (GOAL y (EVENTUALLY p))) ∧
(PREFER x p ¬p) ∧
(P_R_GOAL x p (GOAL y (EVENTUALLY p)))

Différents types de coopérations :

- **Coopération accidentelle**
- **Coopération intentionnelle unilatérale**
- **Coopération mutuelle**
- **Adoption coopérative**

L'organisation

- **organisation (Morin 77) = agencement de relations entre composants :**
  - qui produit une unité, dotée de qualités inconnues au niveau des composants
  - qui lie de façon Inter-relationnelle des éléments qui dès lors deviennent les composants d'un tout
  - qui assure solidarité et solidité relative
- **organisation (Fox 81) = schéma** décrivant comment les membres de l'organisation sont en relation et interagissent afin d'atteindre un but commun : les sorties de l'organisation
- **organisation (Malone 87) = structure de coordination et de communication** comprenant un ensemble d'acteurs
- **organisation du travail/organisation sociale (Rasmussen 91):**
  - l'allocation d'une tâche particulière se développera pour chaque situation, guidée par les compétences des acteurs et la "technologie du domaine de travail" et déterminera le contenu de la communication inter-agents,
  - l'interaction sociale entre les agents dépend de la forme de la communication, qui à son tour dépend de la stratégie de la coordination adoptée

Auto-organisation

- **modification de la topologie du groupe décidée de façon autonome**
- **adaptation à l'environnement :**
  - par spécialisation de fonctions (apprentissage)
  - par modification de la topologie du groupe
- **connaissance de l'organisation sociale :**
  - par observateur externe: seulement des relations inter-individuelles pour les agents
  - par l'agent de la société: représentation de la collectivité permettant de la structurer

Émergence

Étude des conditions "d'émergence" d'un comportement intelligent :

- issu de **l'agrégation d'entités** plus simples
- **non directement "programmé"** dans chacune d'elles.

Introspection

- **capacité à raisonner** sur ce qu'on **fait** et sur ce que l'on **est**
- **caractérisation des espèces (groupes)**
- **tous les agents appartiennent à la même espèce**
  - pas de nécessité de représentation de soi,
  - chaque entité est seulement un individu
- **existence d'espèces (agents hétérogènes)**
  - propriétés par défaut pour de nouveaux individus
  - raisonnement sur des individus non-connus
  - nécessité de communication par diffusion totale ou relaxation
  - caractérisation (critères fonctionnels, structurels,...
- **construire une image de soi (conscience)**
- **raisonner sur les relations avec les autres**

Auto-reproduction

Production d'une nouvelle unité semblable :

- copie conforme ?
- auto-organisation du nouvel agent
- choix des caractères à dupliquer (espèce)

Société insuffisamment développée pour atteindre ses objectifs :

- création spontanée de nouveaux agents
- complexification de la société
- adaptation à l'environnement complémentaire de la redondance initiale

Répartition du travail :

- par distribution de tâches identiques
- par spécialisation fonctionnelle

Contraintes :

- possibilité pour des agents "virtuels"
- difficulté pour des agents réels (robotique)

Perspectives

Autonomie des agents :

- décentralisation du contrôle
- représentation des croyances,
- rationalité, intentionnalité

Approche empirique -> formalisation, théorisation

- méthodologie de conception
- modélisation des communications

Évolution de la société :

- organisation dynamique adaptative
- évolution des compétences des agents
- spécialisation, remplacement