

UQAM

DIC9410 – Présentation du projet de recherche

*Reproduction de mécanismes de "conscience"  
dans un agent logiciel:  
application à un agent d'enseignement intelligent*

**Daniel Dubois**

Laboratoire GDAC, UQAM

19 mai 2005

# *Plan de cette présentation*

- ◆ Présentation sommaire du projet (4 min.)
- ◆ Hypothèses de développement menant aux EIAH actuels (1 min.)
- ◆ Problématique identifiée (1 min.)
- ◆ Mieux situer la conscience pour proposer le projet de recherche (9 min.)
- ◆ Solution proposée (21 min.)
- ◆ Un exemple virtuel (10 min.)
- ◆ Contributions de ma recherche (2 min.)
- ◆ Échéancier prévisionnel (2 min.)

***Note : Cette recherche est un effort collaboratif!***

## *But de cette recherche*

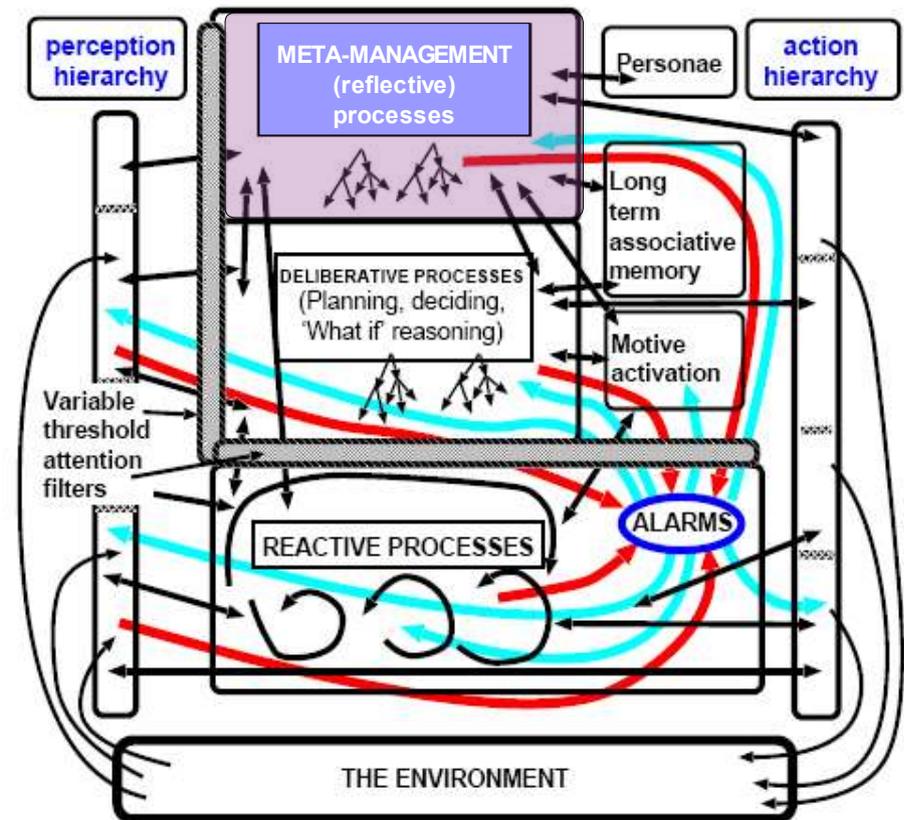
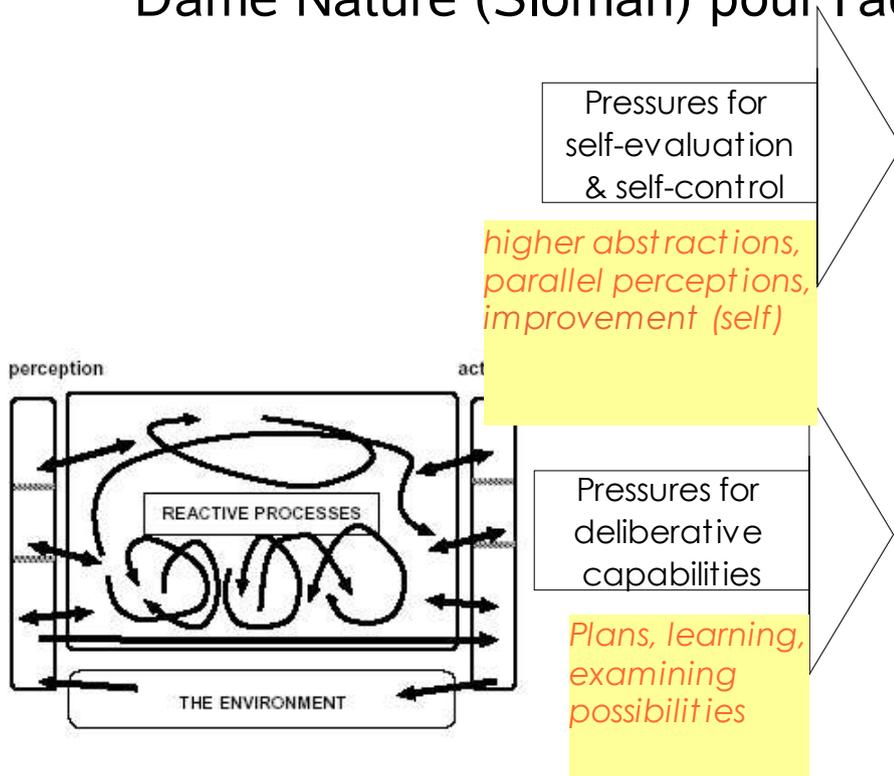
- ◆ Démontrer la capacité supérieure d'adaptabilité que recèlent pour un agent artificiel des mécanismes reproduisant la conscience humaine couplés à un réseau d'actes.

# *Objectifs principaux de la recherche*

- ◆ Poser les jalons principaux d'un agent intelligent pour l'apprentissage humain essentiellement doté de mécanismes reproduisant le cycle cognitif humain et la **conscience**.
- ◆ Implanter les capacités de l'agent (pédagogiques et autres) principalement dans un **réseau d'actes** intimement couplé aux mécanismes de la "conscience" artificielle.
- ◆ Fonder le réseau des actes sur une **théorie pédagogique** -- peut-être sur plusieurs.

# Pourquoi un fondement aussi étrange?

- ◆ La **conscience** est la réponse la plus récente, la plus évoluée de Dame Nature (Sloman) pour l'adaptabilité.



Source : Sloman, Aaron et Ron Chrisley (2004). "Virtual Machines and Consciousness". *In Journal of Consciousness Studies*, 10, no 4-5.

*Les agents artificiels  
doivent aussi s'adapter*

# *Ces EIAH actuels: Nécessité de vraiment considérer l'apprenant*

- ◆ Selon les théories pédagogiques récentes :
  - ◆ Pour améliorer l'apprentissage, l'EIAH doit s'adapter à l'apprenant et à l'ensemble de son contexte.
  - ◆ L'apprenant n'est pas une collection de règles bien structurées dans laquelle on colle de nouvelles informations.
    - ◆ Il est sensible à son environnement, présente une personnalité teintée par son histoire, vit des états d'âme, etc. et ses dispositions à l'apprentissage peuvent fluctuer dans le temps.
    - ◆ Les capacités d'apprentissage et les préférences varient d'un apprenant à l'autre.
- ◆ Trois nécessités pour l'adaptation :
  - ◆ 1- Préparer l'adaptation :  
séparer les champs et les modéliser (MA, MD, expertise)
  - ◆ 2- Faciliter la communication avec l'apprenant :  
personnage animé, langue naturelle, interprétation sémantique
  - ◆ 3- Adapter la présentation :  
contenu, style, mise en contexte

*Ces EIAH actuels:*

## *La problématique repérée : COMPLEXITÉ*

- ◆ Selon les théories pédagogiques les plus récentes, l'efficacité de l'apprentissage repose (pour une bonne part) sur la capacité du EIAH de considérer un grand nombre de facteurs (incluant les émotions)
- ◆ La planification des interactions entre ces facteurs excède la capacité des concepteurs et/ou celle des machines (explosion combinatoire).
- ◆ L'objectif à viser :  
**Gérer la complexité pour adapter l'agent d'enseignement.**

# *Piste pour la solution : Comment l'humain s'adapte-t-il?*

- ◆ Une piste : comment agissons-nous de manière appropriée?
  - ◆ Nous tenons compte d'un maximum d'informations pertinentes
    - ◆ Nous filtrons donc (temporairement) celles non pertinentes à la situation
    - ◆ Nous ordonnons certaines informations par priorité
  - ◆ Nous analysons nos propres mécanismes, notre propre pensée (réflexion)
  - ◆ Nous retenons les leçons (apprentissage, perfectionnement)
  - ◆ Nous parvenons à faire cela en temps réel!
- ◆ Quel mécanisme nous permet ainsi de faire mieux (globalement) que le reste des créatures?
  - ◆ le mot commence par "con" (...prochaine page)

*C'adaptabilité humaine*  
*La conscience*

- ◆ Surpris?
  - ◆ la conscience n'est pas un algorithme d'analyse ou de résolution de problèmes;
  - ◆ la conscience n'est pas une réserve d'adrénaline aidant à faire face aux situations dangereuses.
- ◆ La conscience désigne une foule de mécanismes permettant l'adaptabilité (efficente, bien entendu).
- ◆ La multiplicité des acceptions du mot est à la source de notre méfiance envers cette chose.
  - ◆ on la réduit parfois à "conscience professionnelle", "éthique", "morale".
  - ◆ Elle inclut ces idées et en contient d'autres encore (...prochaine page)

# *Ces descriptions scientifiques de la conscience*

- ◆ La science (philosophie et psychologie) précise mieux cette grande valise. Selon Ned Block, elle regroupe plusieurs mécanismes :
  - ◆ la conscience **de supervision**
    - ◆ les processus qui supervisent nos sens et "nous" tiennent informé des perceptions
  - ◆ conscience **d'accès** (*qui donne accès*)
    - ◆ rend disponible à l'ensemble de nos sous-systèmes les représentations de nos expériences;
    - ◆ ceci permet le recrutement (indirect) des ressources inconscientes autrement inaccessibles
  - ◆ La conscience **du Soi** (nous existons en tant qu'entité séparée du reste du monde)
  - ◆ La conscience **phénoménale** (la qualité des expériences, *ce-que-ça-fait-d'être*, la réalité difficile à mettre en mots)

*C'adaptabilité humaine : la conscience*  
*N'est-ce pas bien abstrait?*

- ◆ Des hypothèses sur son fonctionnement biologique confirment la concrétisent de plus en plus.
- ◆ Baars (1988, 1997, 2002) organise de multiples théories en un cadre général.
  - ◆ Le système nerveux : une collection de réseaux spécialisés distribués (processeurs simples et inconscients), associés à un espace global (une sorte de mémoire éphémère dont le contenu est diffusé aux processeurs inconscients).
  - ◆ L'espace global sert aussi à intégrer les divers réseaux d'entrée compétitifs et collaborateurs.
  - ◆ Certains réseaux, bien qu'inconscients, influencent le contenu conscient (*contextes*)
- ◆ Crick et Koch (1995) confirment et incarnent cette théorie dans la synchronie des oscillations de divers modules du cerveau.

*C'adaptabilité humaine : la conscience*

## *N'est-ce pas encore bien abstrait?*

- ◆ Certains mécanismes de l'esprit humain, incluant ceux de la conscience, ont déjà été intégrés dans les ordinateurs :
  - ◆ processus de supervision, sélection des informations
  - ◆ processus automatiques compilés, parallélisme
- ◆ Il faut dire que plusieurs aspects de la conscience peuvent intéresser les architectes de systèmes et les chercheurs en IA (...proch. page)

*C'adaptabilité humaine : Pourquoi un fondement aussi étrange?*

## *Quelques fonctions attrayantes*

- ◆ La conscience permet de relier le système dans son ensemble
- ◆ Une connexion pénétrante
  - ◆ quel que soit son emplacement, l'information se fait connaître à l'ensemble du reste du système.
    - ◆ ex.: un estomac creux, une coupure, un résultat inattendu, un changement, etc.
- ◆ Une connexion flexible
  - ◆ elle crée une connexion temporaire entre systèmes non connectés
- ◆ Son fonctionnement résout la difficulté rattachée à la pertinence
  - ◆ les ressources inconscientes réagissent et se proposent par elles-mêmes
- ◆ Elle unifie diverses sources et modalités sensorielles
- ◆ Elle amène une utilisation efficace des ressources
  - ◆ un seul élément conscient à la fois → sélection de la priorité pour le traitement explicite
- ◆ La référence au Soi oriente et uniformise les décisions

# *La proposition du projet*

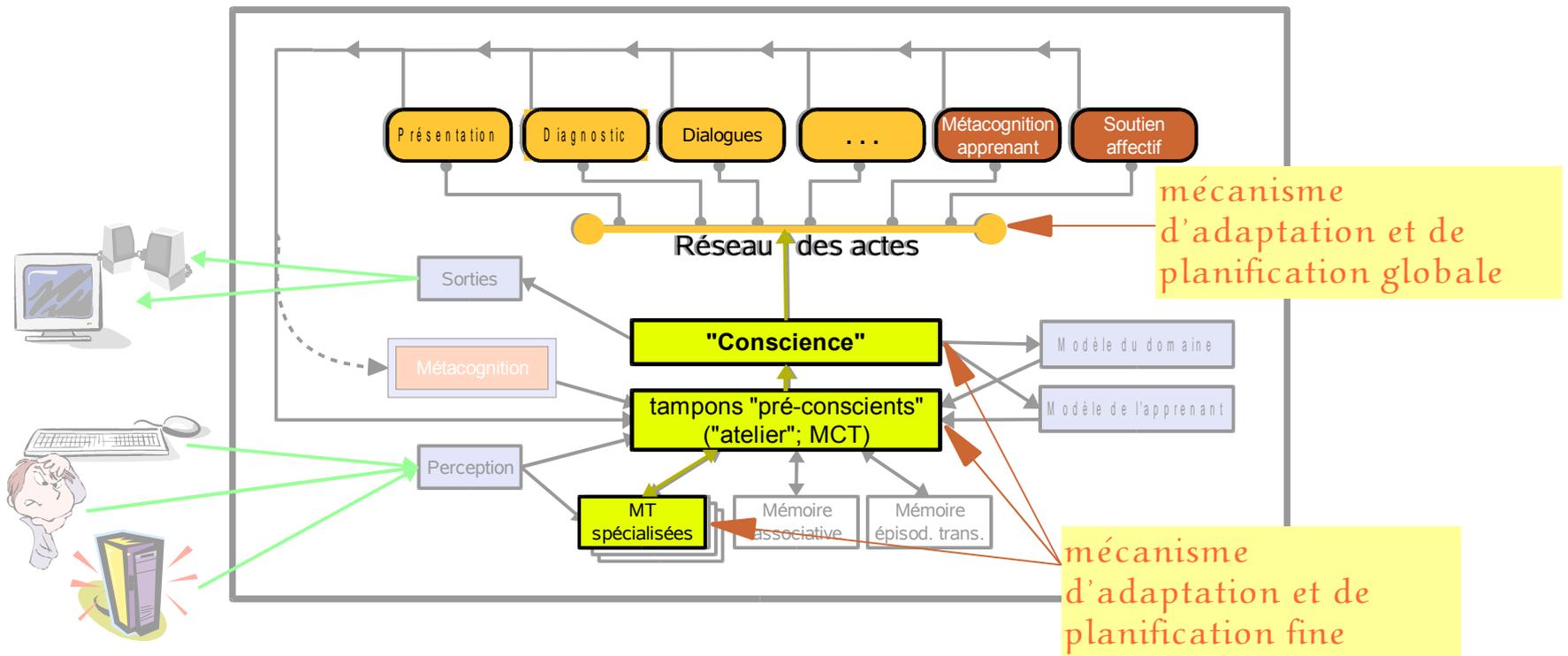
# *Le projet de recherche*

- ◆ Construire un ACA sur la base de l'agent IDA élaboré par le professeur Stanley Franklin.
- ◆ IDA : *Intelligent Distribution Agent* – un agent d'assignation "conscient" développé pour la Marine des États unis [Franklin et al. 1998]
  - ◆ Réalise une mise en oeuvre **fonctionnelle** psychologiquement plausible de la conscience
  - ◆ Fait usage d'un réseau d'actes couplé à cette "conscience"
  - ◆ Intègre une architecture cognitive complète
- ◆ Adapter ses éléments et ajouter les aspects manquants

...Vous brûlez de rencontrer IDA dans son incarnation pédagogique?



# L'architecture du ACA



# *Le réseau des actes*

## *(avec quelques extensions de Franklin)*

- ◆ Un graphe orienté **d'enchaînements** d'actes (plans partiels);
- ◆ un noeud de but chapeaute chaque enchaînement ("signal d'achèvement" et lien avec les *pulsions*);
- ◆ structure des noeuds d'acte :
  - ◆ une liste de préconditions
  - ◆ l'acte à exécuter (spécification de haut niveau)
  - ◆ deux listes des changements attendus
- ◆ **Les actes sont réalisés par des *microprocessus* (*codelets*)**
- ◆ Ils implantent essentiellement les *processeurs simples* de Baars
- ◆ Mais il y a plus : une circulation d'énergie (page suivante)

# *C'IA Nouvel-Âge : des infusions énergétiques*

- ◆ De l'énergie coule dans ce réseau
  - ◆ en provenance du "sommet" : les *pulsions* de l'agent
    - ◆ pulsions pédagogiques ("professionnelles")
    - ◆ pulsions générales (maintenance, interactions générales, survie, etc.)
  - ◆ en provenance du "bas" (appliquée directement aux noeuds) : infusée par les divers événements et états
- ◆ Pourquoi? *page suivante...*

## *Elle mène un peu, beaucoup, passionnément...*

- ◆ Le niveau énergétique d'un noeud (acte ou but) :
  - ◆ *indique à tout moment la pertinence d'un noeud par rapport à la situation globale.*
    - ◆ Les noeuds accumulent l'énergie reçue de toutes les sources (pulsions de l'agent, états et événements internes et externes, énergie reçue des autres noeuds d'acte).
    - ◆ L'énergie constitue une transmutation en une "monnaie commune" de la valeur d'un noeud pour chaque facteur.
    - ◆ Il montre combien un acte convient à plusieurs facteurs, ou à des facteurs importants.
- ◆ L'écoulement d'énergie est primordial pour la réactivité
  - ◆ Il correspond à une analyse perpétuelle de la situation globale.
    - ◆ Ainsi, chaque portion du réseau est constamment à jour avec la situation, et en préparation pour prendre la relève.
- ◆ L'écoulement d'énergie est primordial pour la planification
  - ◆ Les sources d'énergie s'adaptent ou apparaissent au fil de l'évolution de la situation.
  - ◆ L'accumulation d'énergie dans les enchaînements favorisent la complétion d'un plan partiel (une fois en marche).

## *Le réseau et la conscience*

- ◆ Le RA et la "conscience" se complètent.
- ◆ La "conscience" repère, recueille et unifie les diverses informations, effectue une sélection et publie la plus importante coalition.
  - ◆ Cela alimente le réseau des actes en information de qualité.
  - ◆ Cela informe aussi les autres sous-systèmes afin qu'ils puissent réagir et fournir informations pertinentes ou microprocessus dans le but d'aider à traiter la situation
    - ◆ mémoires et MA : informations appropriées à propos de l'apprenant
    - ◆ MD : sujets candidats
    - ◆ RA : microprocessus qui vont sélectionner le prochain concept pour le plan d'exécution, etc.
- ◆ Plusieurs mécanismes de "conscience"
- ◆ Cela réalise la connexion pénétrante et flexible.

*Un peu de sentiments,  
que diable!*



# *Sentimentale, la machine?*

- ◆ Sentimentale? Du moins, douée de "sentiments" et d'"émotions"
  - ◆ pour mieux reproduire la cognition humaine (travaux en cours de Franklin) et/ou améliorer le contact avec l'utilisateur et l'adaptation des présentations.
  - ◆ *Encore seulement dans le modèle conceptuel d'IDA*
- ◆ Émotions et sentiments poussent à l'action
  - ◆ Ils sont un moyen d'évaluation rapide de la situation
    - ◆ Sentiments humains : faim, soif, douleur, fatigue, dépression, etc.
      - ◆ Chez l'agent IDA : ses *pulsions*
    - ◆ Émotions : sentiments avec contenu cognitif
      - ◆ Chez l'agent IDA : Les états affectifs associés à une perception ou à un état interne
- ◆ Les émotions fortes associées aux perception ou aux états internes augmentent pour une coalition les probabilités d'être sélectionnée.

Source : Franklin, Stan et Lee McCauley (2004). "Feelings and Emotions as Motivators and Learning Facilitators", *In Architectures for Modeling Emotions. AAI Spring Symposia Technical Series Technical Reports SS-04-02.*

# *Machinal, le sentiment?*

- ◆ L'agent peut néanmoins traiter les émotions de l'apprenant.
  - ◆ Au même titre que tout *état* perçu
  - ◆ Quoique "machinale", la considération de ce paramètre enrichit l'évaluation de la situation et peut renforcer la proximité à l'utilisateur
  - ◆ Plan actuel : utiliser un agent externe pour la perception des états affectifs de l'utilisateur (ex.: Émilie-2 du GDAC, en développement).

# *La pédagogie dans le projet*



*Les aspects pédagogiques du projet :*

## *Les capacités de l'ACA*

- ◆ L'architecture proposée peut incorporer de nombreuses fonctions et capacités pédagogiques:
  - ◆ Enseigner avec une panoplie de méthodes (en spirale, en dialogue socratique, indices, démonstrations, exemples, etc.);
  - ◆ Adaptation et soumission d'exercices;
  - ◆ Évaluation des compétences (pré- et post-présentation) / Diagnostic des difficultés / Remédiation
  - ◆ Création de dialogues
  - ◆ Support affectif
  - ◆ Stimulation de la métacognition de l'apprenant
  - ◆ Et plus!!!
- ◆ Seule une fraction de celle-ci apparaîtront dans le prototype!!!

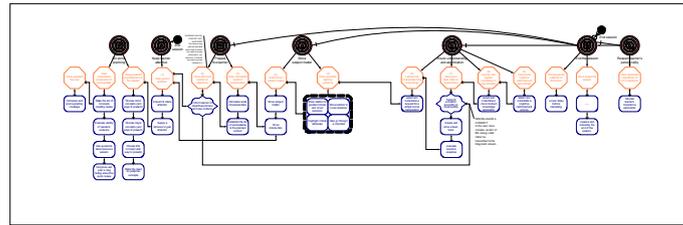
*Les aspects pédagogiques du projet :*

## *Le réseau d'actes pédagogiques*

- ◆ Un réseau d'actes initialement basé sur les théories de Gagné
- ◆ Les neuf événements de l'instruction et les conditions de l'apprentissage présentent une structure claire et hautement compatible à un RA
  - ◆ les événements de l'instruction se transforment en buts (au sommet d'une chaîne d'actes)
  - ◆ les conditions de l'apprentissage indiquent des préconditions aux noeuds d'actes
- ◆ Des vides peuvent être comblés par :
  - ◆ des théories compatibles  
(ex.: Selon Reigeluth : Component Display Theory de Merrill)
  - ◆ des études de vocabulaire et des taxonomies pédagogiques  
(ex.: travaux de Tom Murray (EON), GTE de Van Marck's, etc.)

# *Un RAP simple et entraînant*

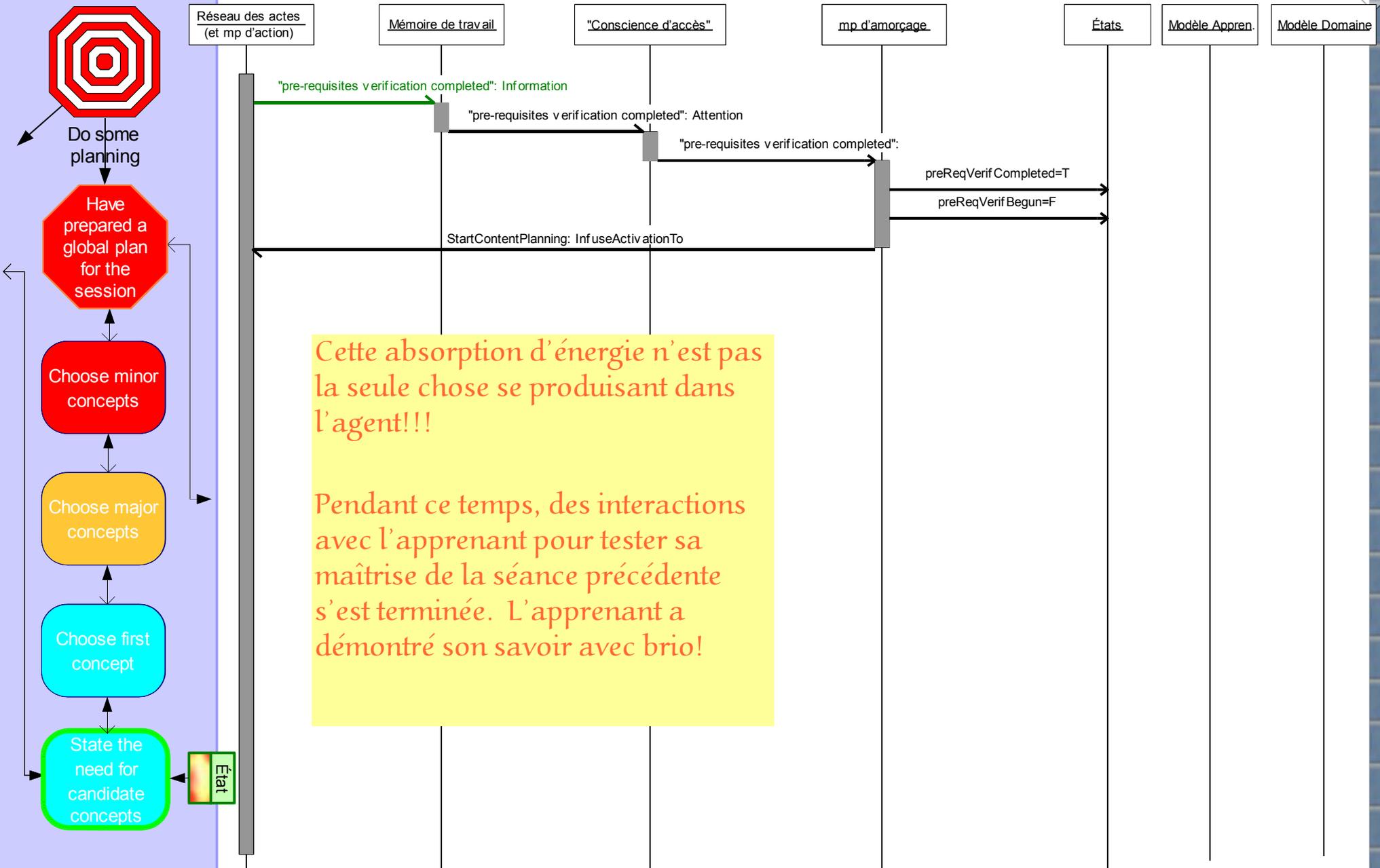
*--- un bref coup d'oeil ---*



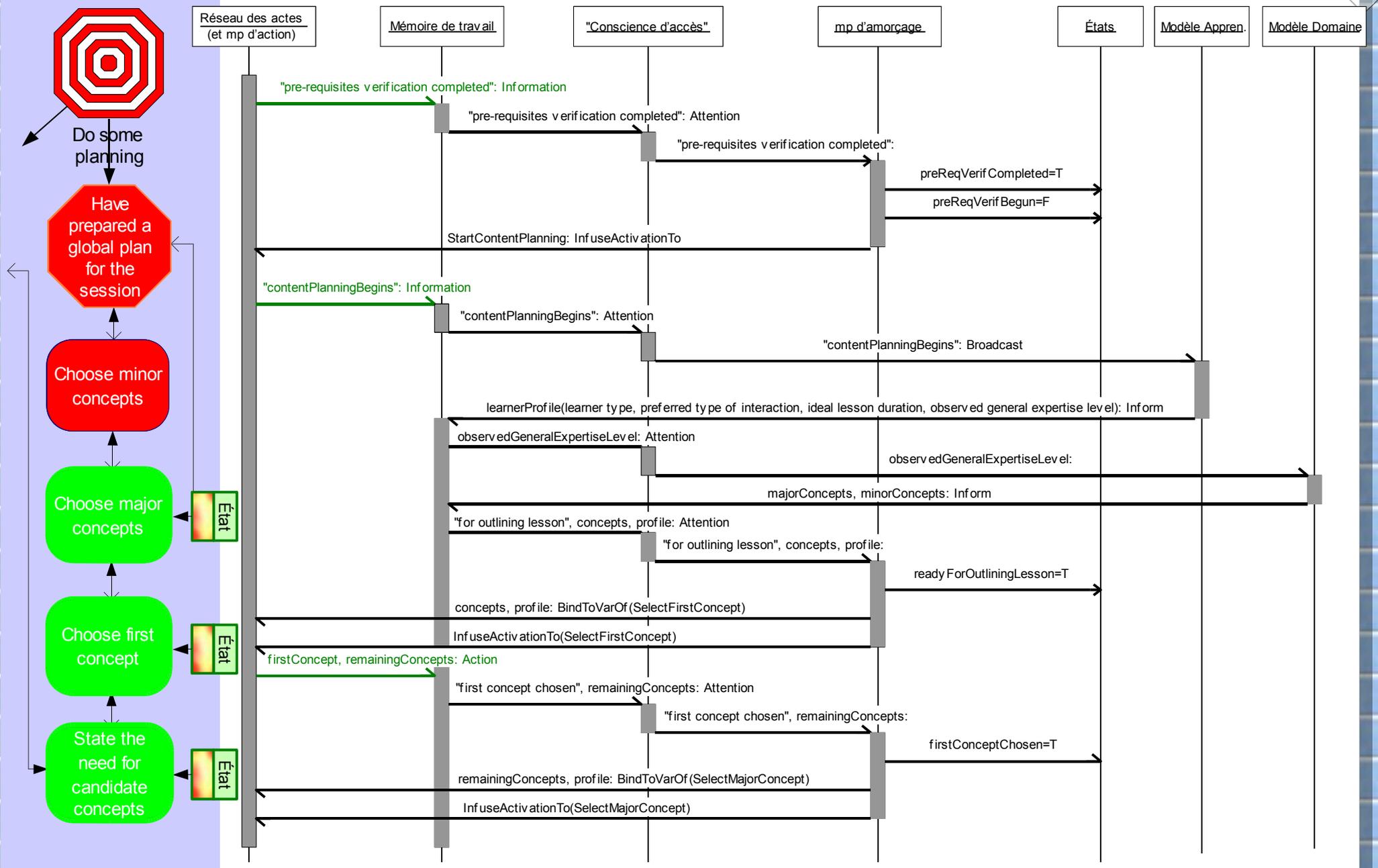
# *Les grands principes de la collaboration RA — mécanismes de la "conscience"*

- ◆ Des plans de présentation spécifiques (détaillés et adaptés) s'élaborent par la collaboration de la "conscience d'accès" et du RA.
  - ◆ Les microprocessus d'action et autres déposent des informations dans l'espace global;
  - ◆ cela est repéré par quelques microprocessus de l'attention, et alors...
  - ◆ la "conscience d'accès" informe tous les sous-systèmes de l'état des choses et des besoins prioritaires.

# Construction interactive d'un plan



# Construction interactive d'un plan



*Quelques derniers points  
généraux*



## *Contribution de cette recherche*

- ◆ Élaboration d'un STI fortement inspiré de la cognition humaine
  - ◆ amélioration de la proximité avec l'utilisateur
- ◆ Application de mécanismes reproduisant la conscience humaine à un domaine aussi complexe que l'enseignement
  - ◆ amélioration de la réactivité
  - ◆ amélioration de la qualité de la planification (planification holistique)
- ◆ Production d'un agent pouvant éventuellement servir à l'exploration de la conscience.

# Échéancier

- |  |           |
|--|-----------|
| ◆ Analyse et réorg. des progr. d'IDA (Patrick Hohmeyer) :                          | 2 mois    |
| ◆ Démarrage d'un système "simpliste"   | 3 mois    |
| ◆ en versions minimalistes : RA, MA, MD, réseau perceptuel                         |           |
| ◆ Un premier prototype   | 5,5 mois  |
| ◆ versions plus riches : <b>RA</b> , MD, rés. percep.(?) + Proc. métacogn. simples |           |
| ◆ Évaluation des performances / réaction de pédagogues                             | 1 mois    |
| ◆ Rédaction  | 2 mois    |
|  | <hr/>     |
|  | 13,5 mois |
| ◆ <b>Suite de cette recherche :</b>  |           |
| ◆ Lien avec une ontologie pédagogique  |           |
| ◆ Amélioration des modules initiaux  |           |
| ◆ Ajout d'autres fonctions : diagnostic, remédiation, autres modes de présentation |           |

# *Sommaire*

- ◆ Un agent pour l'apprentissage humain doté de :
  - ◆ capacité de contextualisation élevée
    - ◆ capacité à considérer un grand nombre de facteurs (dans son adaptation et sa planification)
    - ◆ incluant les émotions de l'apprenant, ainsi que celles de l'agent
  - ◆ bonne séparation des domaines (MA, MD, RA)
- ◆ Une architecture qui reproduit l'esprit et la conscience humains
  - ◆ incluant les sentiments et émotions de l'agent (meilleur contact avec l'apprenant)
  - ◆ incluant la personnalité (dans ses pulsions et ses mémoires, ou dans un module)
- ◆ Une architecture aisée à enrichir et maintenir
  - ◆ possibilité d'intégrer naturellement des fonctions additionnelles
- ◆ Une solide base de départ : IDA

*À votre tour!*