

# Conscience artificielle

Guillaume Tisserant  
Guillaume Maurin  
Ndongo Wade  
Anthony Willemot

2010



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Présentation du rapport . . . . .	5
1.1.1	Un mot d'introduction . . . . .	5
1.1.2	Présentation du projet . . . . .	5
1.2	Les participants . . . . .	5
1.2.1	remerciement . . . . .	5
1.2.2	répartition des tâches . . . . .	5
<b>2</b>	<b>Qu'est ce que la conscience ?</b>	<b>7</b>
2.1	Définition de la conscience : . . . . .	7
2.1.1	Une définition suffisante ? . . . . .	7
2.1.2	La conscience, un atout d'évolution ? . . . . .	7
2.1.3	La conscience complète comme propriété de l'Homme . . . . .	8
2.1.4	Illustration d'un processus récursif . . . . .	8
2.2	La représentation mentale . . . . .	9
<b>3</b>	<b>Les différentes modélisations existantes</b>	<b>11</b>
3.1	La conscience artificielle selon Ray Kurzweil. . . . .	11
3.2	Le modèle d'Alain Cardon . . . . .	11
3.3	Gérard SABAH et conscience artificielle . . . . .	12
<b>4</b>	<b>Notre modèle de représentation de la Conscience</b>	<b>15</b>
4.1	Présentation du modèle . . . . .	15
4.1.1	Pourquoi un nouveau modèle ? . . . . .	15
4.1.2	Version simplifiée du modèle . . . . .	16
4.2	L'inconscient dans notre modèle . . . . .	16
4.2.1	Le cerveau reptilien . . . . .	16
4.2.2	Le cerveau de la mémoire . . . . .	17
4.3	Le Néocortex . . . . .	17
4.3.1	Le préconscient . . . . .	17
4.3.2	Le conscient . . . . .	19
4.3.3	La synthèse . . . . .	19
4.4	Version du modèle contenant des émotions . . . . .	20
4.4.1	La version finale du modèle . . . . .	20
4.4.2	Un moyen de communiquer entre les zones du cerveau. . . . .	20
4.4.3	Fonctionnement théorique. . . . .	20
4.4.4	Fonctionnement algorithmique . . . . .	21
4.5	Pertinence de notre modèle . . . . .	21
4.5.1	Notre modèle, une représentation d'une conscience humaine . . . . .	21
4.5.2	Le modèle et les phénomènes psychanalytiques . . . . .	22
4.5.3	Remarques diverses sur le modèle . . . . .	22
4.5.4	Vers une conscience artificielle ? . . . . .	23
<b>5</b>	<b>L'utilité d'une conscience artificielle</b>	<b>25</b>
5.1	En IA "classique" . . . . .	25
5.2	En paradigme agent . . . . .	25
5.3	En robotique . . . . .	26
5.4	Au-delà de la recherche . . . . .	26



# Chapitre 1

## Introduction

### 1.1 Présentation du rapport

#### 1.1.1 Un mot d'introduction

Le thème de la conscience artificielle agite les esprits des romanciers et des scientifiques depuis l'apparition des premières machines capables de calculer. C'est un sujet qui relie des compétences à la fois de neurologues, d'informaticiens mais aussi de psychologues, d'anthropologues et même de philosophes. Pour ce mémoire, nous ne nous sommes pas placés en tant qu'informaticiens essayant de modéliser une conscience sur nos machines, mais nous avons essayé d'avoir une vision beaucoup plus large du problème, en partant du point de vue de l'être humain plutôt que de celui de l'ordinateur.

#### 1.1.2 Présentation du projet

La création d'une conscience artificielle est quelque chose qui pourraient ouvrir énormément de nouvelles portes en informatique, donnant à des ordinateurs des capacités qu'ils ne peuvent pas avoir dans l'état actuel de l'informatique, et qui sont pour l'instant réservés au monde animal, voir à l'être humain. Ce rapport a pour but de montrer comment une conscience artificielle pourrait être créée. Nous commencerons par donner une définition de la conscience puis nous ferons un état de l'art sur les modèles de conscience artificielle. Ensuite nous proposerons le notre en expliquant ses avantages par rapport à ceux proposés précédemment. Enfin nous expliquerons ce que pourrait offrir cette conscience artificielle.

### 1.2 Les participants

#### 1.2.1 remerciement

Nous remercions Violaine Prince, qui nous a suivi tout au long de la création de ce rapport, de nos premières réflexions à la rédaction finale, nous donnant de nombreux conseils et précisions. Nous tenons aussi à remercier Gina Devau, qui a accepté de relire notre rapport et de nous donner son point de vue de Neurobiologiste.

#### 1.2.2 répartition des tâches

Le projet a été initié et dirigé par Guillaume Tisserant qui a proposé son modèle de conscience artificielle et a fait une grande partie du travail de recherche, secondé par Guillaume Maurin qui s'est concentré sur la définition de la conscience et la question de savoir si le modèle en question nous menait bien vers une conscience ou non. Ndongo Wade s'est chargé de l'état de l'art des modèles existants et Anthony Willemot a travaillé sur l'intérêt de la création d'une conscience artificielle.



## Chapitre 2

# Qu'est ce que la conscience ?

### 2.1 Définition de la conscience :

Il est important de commencer par se mettre d'accord sur une définition de la conscience. C'est en effet un terme dont chacun peut avoir une interprétation propre. La conscience peut simplement se définir par la connaissance qu'un être vivant a de lui-même et de son environnement. Par connaissance, on entend l'image mentale qu'il s'en fait, qu'elle soit proche de la réalité ou non.

#### 2.1.1 Une définition suffisante ?

Dans cette vision de la conscience, on pourrait dire que n'importe quel être vivant est conscient du moment qu'il se représente dans son esprit les choses de manière abstraite. Mais il ne faut pas confondre conscience et mémoire, la conscience présuppose une capacité de "gestion" des informations contenues dans la mémoire, soit pour créer des nouveaux concepts, soit pour modifier ceux existants par pure introspection. Il y a donc une différence entre connaissance et conscience. On pourrait parler de raisonnement, mais il ne s'agit pas simplement d'un enchaînement de déductions logiques tel que pourrait le faire un système expert dans une machine. Là où le raisonnement se cantonnerait à combiner des données brutes captées par les percepts pour créer une nouvelle donnée, telle une fonction mathématique prenant des paramètres en entrée et en sortie, la conscience serait un méta-système capable de directement "raisonner sur le raisonnement". L'individu est capable de penser sur sa propre pensée et de se représenter ses propres représentations. On appelle également cela "conscience introspective".

#### 2.1.2 La conscience, un atout d'évolution ?

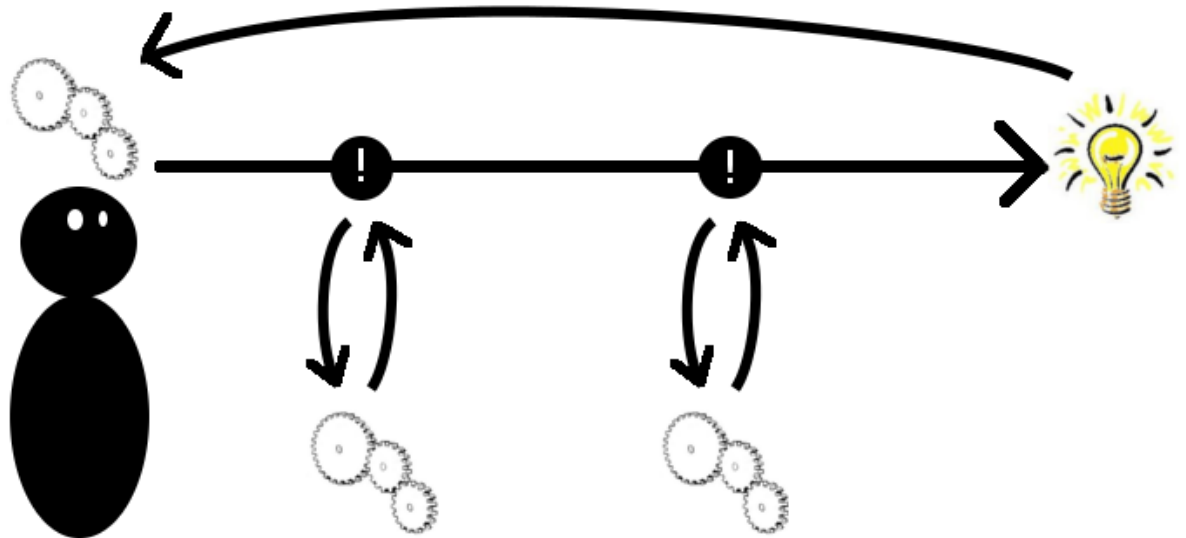
A partir du moment où un tel niveau de conscience de soi est atteint, le "système" peut évoluer de manière beaucoup plus efficace et s'adapter à toute nouvelle situation. On voit rapidement quel a été l'avantage évolutif chez l'Homme. Mais si l'adaptation rapide à un problème naturel est l'avantage le plus visible à être conscient, cela permet aussi une richesse de comportement théoriquement illimitée. Basé sur l'adaptation, l'individu conscient peut donc réagir de façon personnelle et nouvelle à une situation donnée, la réaction n'étant pas forcément provoquée par un besoin direct et de bas niveau mais par toute une série de données et de processus qui auront été plus ou moins modelés par l'individu au cours de sa vie. La réaction en question sera perçue par l'individu lui-même, et par les autres, et sera instantanément analysée pour une remise en question voire une modification. On est au niveau d'apprentissage ultime : il est permanent et ne nécessite pas de répétition d'essais erreurs, et il est en "temps réel". Mieux encore, l'apprentissage peut se faire sans qu'aucun percept extérieur particulier ne vienne stimuler l'individu, une introspection peut être réalisée à un moment ultérieur à l'action, et l'individu peut modifier sa conscience de lui-même et de l'extérieur de manière autonome. Bien sûr, il se servira de tous les percepts qu'il a emmagasinés dans son Histoire, mais leur interprétation est personnelle et peut être faite à n'importe quel moment.

Cependant, l'individu conscient a beau avoir une grande capacité d'interprétation et de réorganisation de son savoir, plus il emmagasine de connaissance plus ce travail de réflexion pourra être fait de manière efficace.

### 2.1.3 La conscience complète comme propriété de l'Homme

Cette définition réserve quelque peu la conscience à l'Homme, puisqu'elle nécessite une mémoire sémantique de haut niveau ainsi qu'une "mémoire de la mémoire" aussi appelée méta-mnèse. Elle permet de se souvenir du déroulement d'une action ou d'un raisonnement afin de pouvoir en comprendre les tenants et aboutissants, et les modifier si besoin. Dans notre exposé nous parlons donc de "conscience humaine" bien que l'on ne doute pas que certains animaux puissent raisonner ou avoir une certaine conscience de leur existence. Il nous a semblé que le seul animal à avoir une conscience "complète", c'est à dire réunissant toutes les caractéristiques supposées nécessaires à notre définition, est l'Humain.

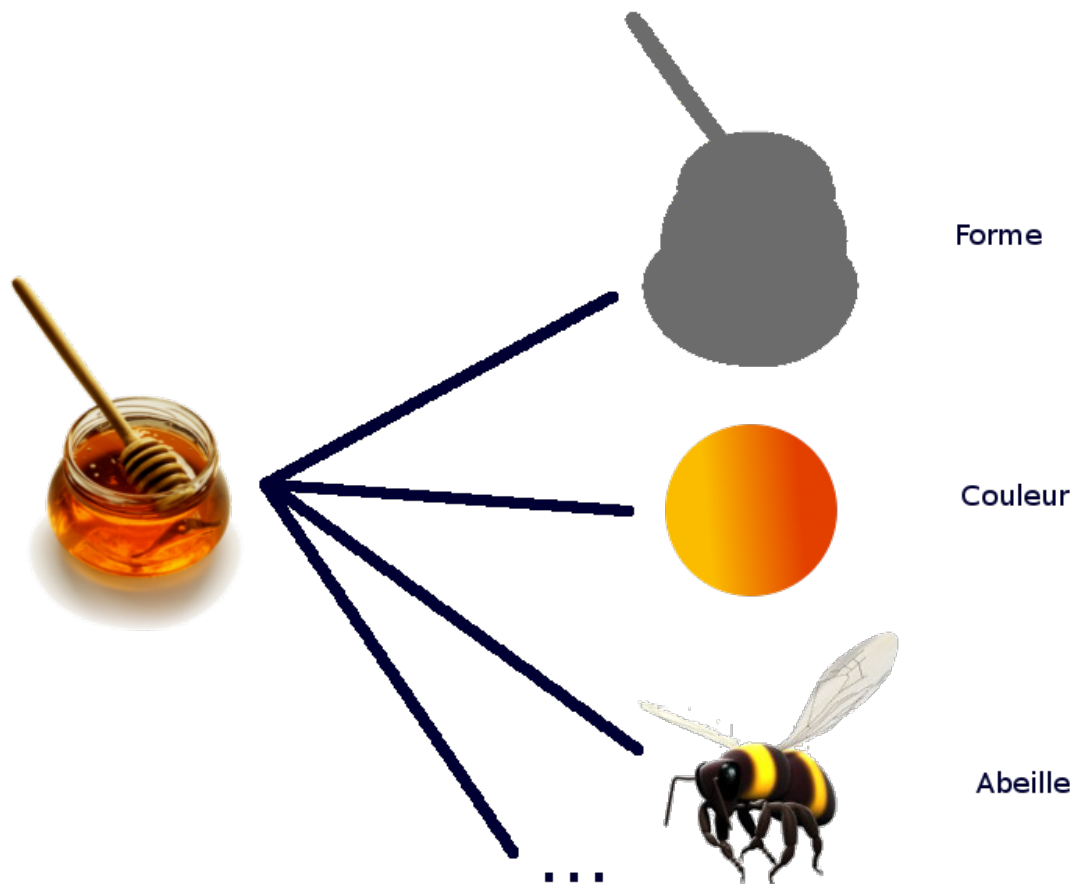
### 2.1.4 Illustration d'un processus récursif



Le schéma ci dessus illustre cette définition où le processus de raisonnement interagit avec lui même pendant son propre déroulement, et utilise ses propres résultats pendant et après la fin supposée du raisonnement. Ce schéma fait penser à celui des circuits neuronaux récurrents ou réverbérants, qui font revenir un signal électrique au neurone émetteur afin qu'il ait un retour immédiat sur le "cheminement" en cours. Il s'agit peut-être du même processus porté au niveau le plus haut, le niveau conscient. Ce type d'idée est assez proche de celles de Lecerf qui propose une vision du fonctionnement de notre cerveau basé sur une double récursivité. [Lec97]



## 2.2 La représentation mentale



Les concepts sémantiques manipulés par la conscience ont une organisation bien particulière dans la mémoire. Sur le dessin ci-dessus on voit que le concept du pot de miel est défini par les liens qui le rapprochent des concepts de la forme du pot, de sa couleur... la somme de l'ensemble de ces associations de concepts qui crée le concept de pot de miel. Cette représentation est inspirée des travaux de Collins et Quillian [CQ95]. Certains concepts sont plus haut niveau que d'autres, mais il existe des concepts suffisamment bas niveau qui sont "primaires" comme le concept d'une couleur ou d'une forme. Pour arriver à visualiser un concept sémantique abstrait, celui-ci doit être relié à d'autres concepts haut niveau, mais ceux-ci sont associés à d'autres concepts qui peuvent être plus bas niveau et ainsi de suite jusqu'aux concepts les plus primaires.

Par exemple, si on prend le concept de "Jalousie" qui n'a pas de représentation physique directe, lorsque l'on va visualiser mentalement ce concept on va sans doute d'abord se représenter deux personnes, par exemple une fille et un garçon, et donc associer le concept de "fille" et de "garçon" au concept de jalousie, puis le concept d'amour, puis le concept d'une troisième personne qui sera le concept de "perturbateur", puis le concept de "frustration" qui lui-même sera associé à d'autres concepts plus primaires qu'on est en mesure de se représenter instantanément.

D'autre part, lorsque l'on rêve, les concepts qui sont reliés entre eux peuvent être très proches voire se confondre. Le scénario d'un rêve se construit par ces associations, par exemple si dans votre rêve vous êtes en train de jouer au ballon avec votre enfant, et que pour vous le concept de ballon se rapproche du football, il est possible que dans l'instant d'après vous soyez sur un stade de football en train de disputer une finale de coupe du monde, alors que cela n'avait pas forcément de rapport direct avec ce que vous faisiez juste avant.

Cette représentation a l'avantage de faire penser à un graphe, ce qui est informatisable. Mais il reste à trouver une structure de graphe simulant ces associations conceptuelles, et surtout un moyen de le remplir avec une grande quantité de concepts, primaires ou abstraits, et faire des liens cohérents entre ces concepts. Un tel graphe permettrait notamment à une Intelligence Artificielle de mieux comprendre les liens qui se font entre les mots dans le langage naturel. Des tentatives de construction d'un graphe de concepts existent déjà et donnent des résultats intéressants. Mais

le meilleur moyen de remplir une telle base de données serait de construire un système autonome capable s'explorer le monde et se créer ses propres concepts et ses propres associations, en se basant évidemment sur une base existante.

## Chapitre 3

# Les différentes modélisations existantes

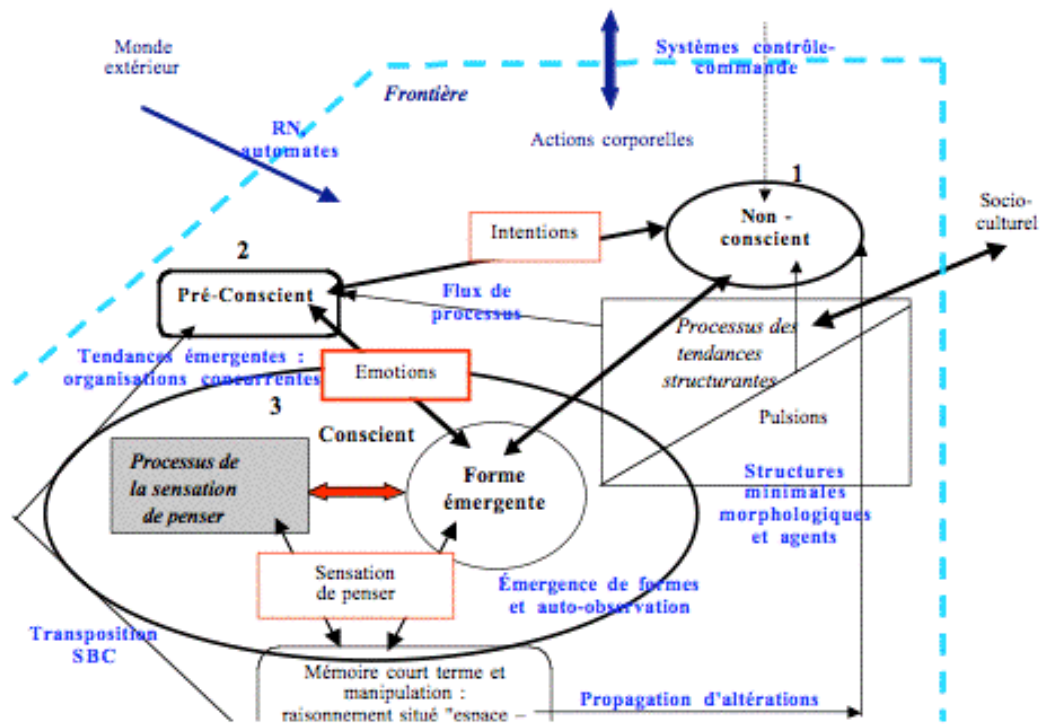
La conscience artificielle est un domaine qui a intéressé bon nombre de scientifiques. Certains ont même proposé des modèles de conscience artificielle. Dans notre exposé nous nous intéressons particulièrement aux travaux de Raymond C. Kurzweil, d’Alain Cardon et de Gérard Sabah.

### 3.1 La conscience artificielle selon Ray Kurzweil.

Ray Kurzweil est un informaticien et futurologue américain né le 12 février 1948. Kurzweil croit à la théorie de la Singularité. Cette théorie stipule qu’avec l’évolution actuelle des technologies de l’information et de la communication on arrivera à un moment où l’intelligence artificielle égalera ou dépassera l’intelligence humaine. Pour lui, un fois ce moment arrivé on pourra télécharger directement le système nerveux de l’humain dans un logiciel. Il prédit que la Singularité arrivera vers les années 2050. [Kur06]

### 3.2 Le modèle d’Alain Cardon

Alain Cardon est un mathématicien français né à 1946 à Paris spécialisé dans le domaine de l’intelligence artificielle. Actuellement, il est membre du laboratoire d’informatique de Paris 6(LIP6). Cardon définit son modèle comme un système psychique artificiel. C’est un système qui se greffe sur les systèmes générateurs présents dans l’informatique actuelle c’est-à-dire de type “entrées traitement-sorties”. Il pense que son système pourra permettre à ces systèmes générateurs de passer d’un non-conscient à un conscient pressenti en passant par un pré-inconscient. Selon lui, ce système est un système évaluateur et superviseur multi-corps présentant en temps réel des représentations avec ou sans inhibition et pathologies, et toujours selon pulsions et interdits, existants, avec évidemment des émotions et le sentiment de soi. Le système est représenté par le schéma suivant :

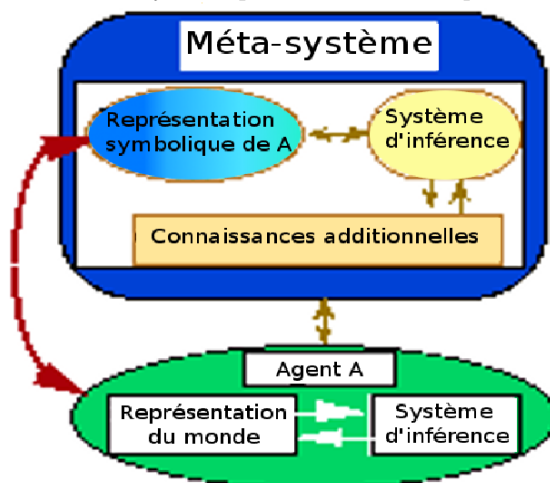


Système psychique artificiel [Car99]

Alain Cardon attend toujours du financement pour réaliser son projet.

### 3.3 Gérard SABAH et conscience artificielle

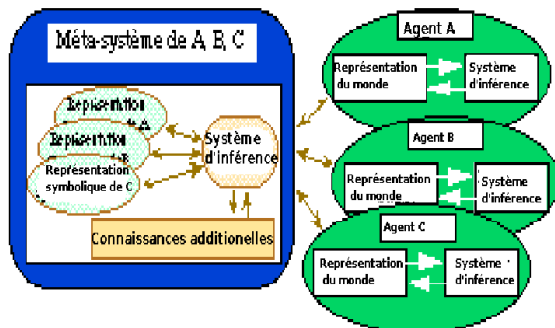
Gerard Sabah est un chercheur français en intelligence artificielle né en 1948 à Paris. Pour modéliser les processus réflexifs, Sabah s'appuie sur le modèle du système réflexif classique de l'intelligence artificielle (fig2) auquel il adjoint deux extensions. Dans ce système l'agent se fait une représentation symbolique de celui avec lequel il va interagir.



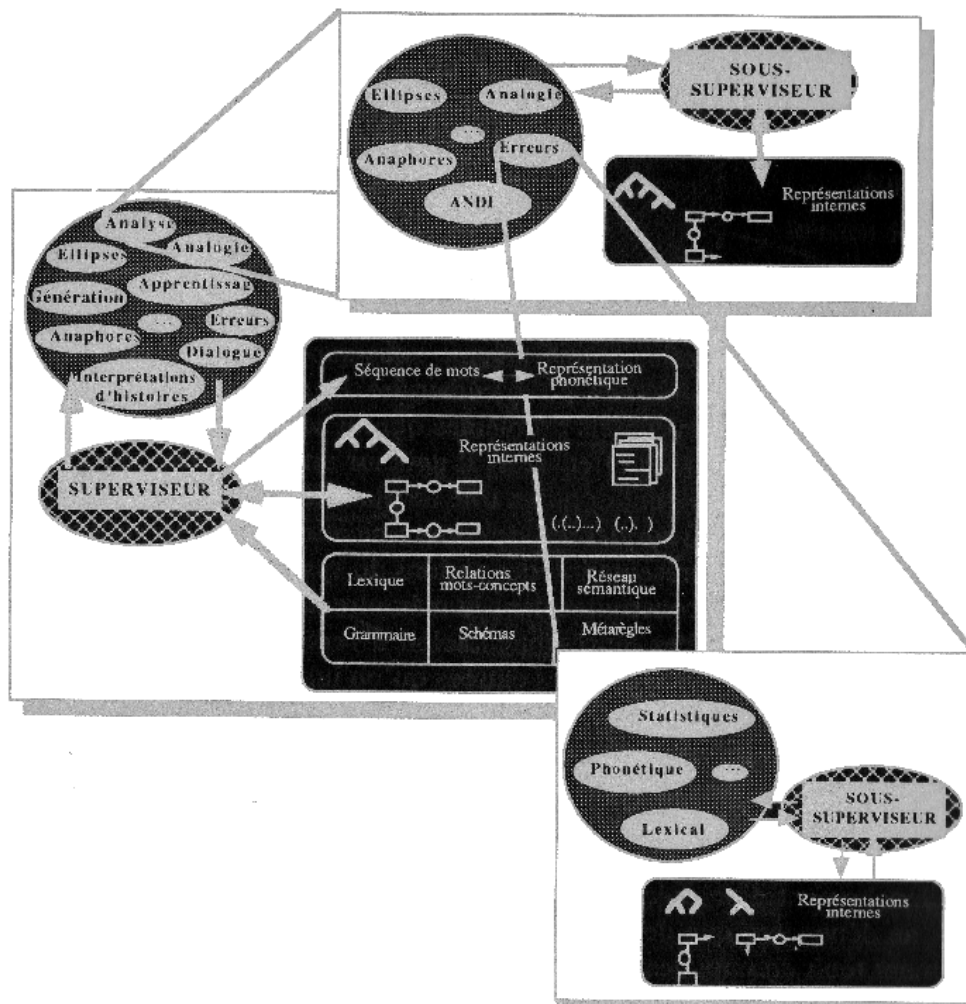
système réflexif

Dans ce modèle chaque agent est contrôlé par un seul méta-système. Sabah pense que ce système est une représentation partielle de la conscience. Donc pour lui, ce modèle peut être étendu de manière qu'un méta-système puisse contrôler plusieurs agents. Ainsi Sabah propose d'étendre ce système par deux ramifications :

a) le fait qu'un méta-système puisse contrôler plusieurs agents  
 b) le fait de considérer ces méta-systèmes comme des agents usuels et leurs appliquer récursivement des méta-représentations. Cette dernière extension fait que chaque agent est contrôlé par un méta-système qui est lui-même un agent et donc est contrôlé par un autre. Ce qui nous amène à un système réflexif. Sabah lui, a pu mettre en œuvre son modèle avec l'implémentation en LISP de CARMEL 1. Voici les schémas des modèles de SABAH [Sab90], [Sab93] :



Un méta-système contrôlant trois agents distincts



Exemple d'une hiérarchie de processus complexes : le premier niveau de contrôle, ayant la tâche de comprendre un texte, active l'analyseur qui, rencontrant un problème imprévu active à son tour simultanément le traitement des erreurs et l'interpréteur d'ellipses.



## Chapitre 4

# Notre modèle de représentation de la Conscience

### 4.1 Présentation du modèle

#### 4.1.1 Pourquoi un nouveau modèle ?

##### **Importance théorique**

Nous nous sommes rendus compte que les modèles précédemment proposés étaient très orientés informatique et ne prenaient que peu en compte les avancées faites dans le domaine de l'étude de l'être humain. Nous avons donc voulu faire notre propre modèle en commençant par faire un schéma le plus proche possible de ce qu'on connaît de l'être Humain, puis ensuite seulement de réfléchir à son implémentation.

##### **Importance pédagogique**

Nous avons rapidement compris que créer un modèle théorique de conscience allait nous permettre de réfléchir par nous même à la faisabilité de la conception d'une conscience, et aux difficultés que devaient rencontrer ceux qui avaient déjà tenté de faire des modèles de conscience artificielle. Concevoir ce modèle nous a permis d'avoir un support pour mieux voir ce qui était faisable ou non avec une conscience artificielle. Ce modèle représente donc notre vision personnelle du fonctionnement que pourrait avoir une conscience artificielle.

##### **Les sources d'inspiration**

Notre modèle utilise beaucoup de travaux de personnes différentes. Nous avons entre autre utilisé le schéma de la mémoire de Tulving [Tul72], le découpage du cerveau de Laborit [Lab76], la vision psychanalytique de notre pensée de Freud [Fre23] et de Lacan, le fonctionnement de la mémoire à long terme de Atkinson et Schiffman [AS68], la représentation des concepts en réseaux sémantiques de Collins et Quillian [CQ95], et la vision réflexive de Sabah [SVV<sup>+</sup>97]. Bien que certaines de ses idées aient dû être adaptées pour une meilleure compatibilité les unes avec les autres et intégrer le modèle, nous avons essayé de rester le plus proche possible des travaux de ces personnes.





schéma essai échec. Les actions entrées dans la mémoire procédurales peuvent se faire de manière inconsciente. Il n'est pas nécessaire de penser consciemment à nos mouvements de jambes pour marcher. Ce morceau de mémoire peut être codé avec des algorithmes d'apprentissage classique tel qu'un réseau de neurones.

## 4.2.2 Le cerveau de la mémoire

### L'analyseur conceptuel

L'analyseur conceptuel est une partie plus complexe que les précédentes. Il nécessite de pouvoir comparer des entrées en dur à une base de données de concepts primaires (la faim, la soif, le plaisir, la fatigue, la joie, une odeur nauséabonde, un bruit violent...) Il faut donc arriver à stocker sous formes de signaux recevables tout les concepts primaires, et pouvoir les comparer aux signaux entrants.

### La mémoire associative

La mémoire associative compare la réception de ces concepts entre eux. Cela permet par exemple à un animal vivant dans un milieu contenant des prédateurs bruyant d'associer le concept de danger au bruit que fait le prédateur.

### L'optimisateur

L'optimisateur a pour vocation d'optimiser un concept qui peut se traduire par le bien-être. Dans l'exemple précédent, il prend en compte que l'approche d'un prédateur réduit le bien-être, et que le fait de courir peut éloigner le prédateur, donc faire remonter le bien être. Il devra donc choisir l'action de fuir.

### Interaction Mémoire Associative/optimisateur

La mémoire associative pourrait être conçue comme une sorte de système expert évolutif, qui analyserait les situations, et essaierait d'en déduire des règles. L'optimisateur serait ensuite chargé de faire le moteur d'inférence et de savoir quel action engager pour optimiser le bien-être. On peut aussi coder la mémoire associative juste comme une mémoire d'association de concepts primaires, et l'optimisateur comme une recherche des concepts les plus rarement associés aux concepts négatifs. Une mémoire associative bien conçue pourrait même faire de l'apprentissage par imitation, en considérant que l'autre est un concept primaire pour les animaux sociaux, et que leur imitation fait monter le concept du bien-être.

## 4.3 Le Néocortex

Le Néocortex est constitué de la préconscience et de la conscience. C'est la partie la plus difficile à réaliser. Elle repousse les limites de nos capacités actuelles en matière algorithmique.

### 4.3.1 Le préconscient

#### l'analyseur conceptuel sémantique

L'analyseur conceptuel sémantique est aussi appelé analyseur conceptuel poussé, pour montrer qu'il manipule des concepts un niveau au dessus de ceux manipulés par l'analyseur conceptuel du cerveau de la mémoire. Il sert à traduire les concepts primaires en concepts complexes stockés dans la mémoire sémantique.

#### la mémoire sémantique

la mémoire sémantique contient une base de données de concepts avancés, construite sous forme de treillis reliant les concepts ensemble [Sow84]. Les concepts présents à l'intérieur ne sont pas écrits en dur mais sont directement créés par le conscient. Elle peut être montré comme un dictionnaire capable de traduire des concepts en en association avec d'autre concepts.

### La mémoire primaire

La mémoire primaire contient un très petit nombre de concepts sémantiques. Elle contient les choses auxquels auquel une conscience est en train de penser. Elle peut être remplis par l'analyseur conceptuel sémantique qui lui envoie les concepts se trouvant dans l'environnement, ou directement par la réflexion consciente qui choisit d'y poser un concept pour réfléchir dessus.

### le problème du boot strap

La mémoire sémantique demande un amorçage. En effet, la mémoire sémantique contient les concepts qui permettent à l'analyseur conceptuel sémantique de traduire les concepts primaires en concept poussés. Mais les concepts entrants dans la base de données de la mémoire sémantique viennent indirectement de l'analyseur. Ainsi, les concepts présents dans la base de données de la mémoire sémantique sont nécessaires à la création de la base de données de la mémoire sémantique.

### Une solution pour le contourner

La mémoire sémantique elle même est construite sur la mémoire associative qui utilise des concepts primaires. On peut donc construire un système qui permet d'associer à des séries de concepts primaires stockés dans la mémoire associative, un concept sémantique qui serait de plus haut niveau. En montant petit à petit en niveau, on pourrait arriver à la mémoire consciente, qui serait sémantiquement forte. Il doit donc y avoir une solution pour passer de la mémoire associative à la mémoire sémantique. Cela va dans le sens de la thèse de Versace, Padovan et Nevers qui défendent l'idée qu'il n'y a qu'un seul type de mémoire interprété selon l'analyse que l'on en fait [VNP02]. Ainsi, on peut considérer qu'avec juste des concepts primaires en boot strap, on peut espérer qu'un état de conscience se développe. L'analyseur doit donc commencer par utiliser des concepts créés par la mémoire associative pour remplir la mémoire sémantique. Celle ci commence donc par contenir des associations de concepts primaires auquel elle donne un sens, avant d'utiliser ces nouveaux concepts pour construire d'autres "mots" dans sa mémoire.

### Gestion de la mémoire sémantique

Les concepts en mémoire primaire doivent être copiés en mémoire sémantique instantanément. La gestion de la mémoire sémantique est aussi complexe. Il faut que tout ce qui est en mémoire primaire y arrive, mais que les choses soient oubliées petit à petit, et que plus elles restent ou repassent en mémoire primaire, moins elles disparaîtront de la mémoire sémantique. Un des problèmes de la mémoire sémantique est qu'elle stocke sous forme de treillis un nombre extrêmement important de concepts, mais qu'il est quand même nécessaire de les retrouver relativement rapidement. L'idée serait alors de faire une mémoire multi dimensionnelle. Le treillis en deux dimensions représente le graphe classique tel qu'il est construit, mais à cela vient s'ajouter une troisième dimension qui est la fréquence d'utilisation d'un concept. Ainsi, à chaque fois qu'un concept arrive en mémoire vive, il monte dans la troisième dimension. L'analyseur conceptuel poussé ayant naturellement tendance à renvoyer les informations de plus haut niveau qu'il connaît, celles ci ont naturellement tendance à remonter dans le graphe. Les liens entre les relations doivent aussi être soumis à ce type d'algorithme : Plus une relation entre deux concepts apparaît, plus elle doit être renforcée, et elle doit par ailleurs être allégée si elle ne revient jamais en mémoire primaire. Ensuite, lors de la recherche d'un concept, que l'on ne peut raccrocher à rien, on commence par explorer les couches hautes du graphe, et on descend petit à petit. Si le concept est rattachable à d'autre concept, il faut faire une recherche en utilisant A\* partant de ce concept, et en donnant comme heuristique de choix la montée dans les concepts vers ceux de plus haut niveau, et l'utilisation en priorité des liens les plus forts. Lorsque l'on a plusieurs concepts de départ, il faut lancer plusieurs A\* en partant de chaque concept, et explorer en priorité les nœuds où plusieurs explorations sont passées. Cela offre deux avantages : Le premier est que les concepts utilisés les plus souvent vont être assez rapides à trouver. Le deuxième avantage est que les concepts utilisés les plus récemment vont automatiquement monter au dessus des autres, et donc les concepts en rapport avec le contexte seront aussi assez rapide à trouver. Le même type d'algorithme peut être utilisé pour la mémoire épisodique. On voit bien à ce moment là, la ressemblance avec la mémoire humaine, où les souvenirs que l'on se remémore le plus souvent, ainsi que les plus récents sont les plus rapides à ramener. Cette façon de fonctionner est en partie inspirée des travaux de Atkinson et Schiffman [AS68] sur la mémoire.

### 4.3.2 Le conscient

#### La mémoire épisodique

La mémoire épisodique ou mémoire autobiographique est une mémoire des situations. Elle sert à se remémorer des situations et des émotions vécus sous formes de concepts sémantiques. Elle doit être codée sous forme de graphes de concepts apparaissant dans la mémoire sémantique. Les concepts stockés viennent à la fois des percepts mais aussi de l'état interne de la personne au moment de l'évènement. La mémoire épisodique se remplit comme la mémoire sémantique. Les événements y arrivent lorsqu'ils passent en mémoire primaire, puis ceux qui y restent "montent" pour être trouvable plus facilement, alors que ceux qui n'y repassent pas se retrouvent petit à petit enterrés sous les autres.

#### La méta Mnèse

La méta Mnèse est la mémoire de la mémoire. Son rôle est de stocker la façon dont les éléments se sont enchaînés dans la mémoire épisodique [PAN01]. Elle peut exister sous forme de datage (ajout d'un concept date dans la mémoire épisodique) et sous forme de liens chargés d'un sens sémantique entre les événements pour retrouver facilement leurs précédents et leurs suivants.

#### Le choix conscient

Le choix conscient (ou réflexion consciente) ressemblerait donc à la base à un optimisateur, mais maniant des concepts de plus hauts niveaux, et de natures différentes. Grâce à la méta mnèse, il peut utiliser des séries de situations, et même imaginer des situations passées, présentes ou futures pour prendre des décisions. Il peut ainsi comparer des situations imaginaires et choisir celle vers laquelle il a envie de tendre. Il peut aussi réfléchir sur des concepts abstraits, comme sur ses réflexions passées. Il peut analyser son propre fonctionnement à partir du moment où il arrive à le retranscrire sous forme de concepts.

### 4.3.3 La synthèse

L'objet représenté par la somme est une simplification du schéma retour. Il représente la transformation de tous les signaux en signaux de bas niveaux, et gère la somme des signaux pour choisir lesquels inhiber et lesquels renvoyer en sortie. Son fonctionnement nécessite d'arriver à combiner les signaux entre eux. Par exemple, si une personne choisit de fumer dans le choix conscient, il doit récupérer la façon d'allumer un briquet dans la mémoire procédurale.



### Changement dans le conscient

Pour l'analyseur conceptuel sémantique, cela va avoir des conséquences plus complexes : En effet, à cause de la très faible capacité de la "mémoire primaire", il est très important pour l'analyseur de n'y envoyer que les concepts les plus utiles. En fonction du contexte émotionnel reçu, il va pouvoir choisir d'envoyer différents concepts reçus, en fonction de si ils sont liés ou non aux émotions reçues. De plus, l'analyseur conceptuel sémantique peut aussi envoyer à la "mémoire primaire" des concepts liés aux émotions subtiles, qui ne peuvent apparaître dans les couches plus basses, principalement des émotions liées au temps (comme la nostalgie). Le changement d'état émotionnel se fera lorsque la mémoire vive récupérera des concepts fortement connotés émotionnellement.

#### 4.4.4 Fonctionnement algorithmique

##### Proposition

L'état émotionnel peut se coder de manière relativement simple : La base algorithmique et l'optimisateur lui envoient directement des émotions, (pour la mémoire primaire, il faudrait en réalité un module qui regarderait en mémoire sémantique à quoi sont reliés les concepts présents en mémoire primaire. Mais cela alourdisait le schéma). L'émotion est codée sous forme de chiffre associé aux émotions primaires. A chaque nouveau signal reçu, le chiffre relié au concept attaché au signal augmente proportionnellement à la force du signal. Le chiffre baisse par contre rapidement lorsque aucun signal pour le maintenir ou l'augmenter n'est reçu.

##### Evolution

Bien que relativement simple, ce modèle devrait permettre de donner des résultats intéressants. En prenant en compte que les émotions peuvent être rattachées de manière complexe à des concepts sémantique, il permet de représenter toutes les émotions possibles dans le conscient. Et le schéma montre bien que les émotions ont des interactions avec tout les endroits du cerveau. Toutefois, il faudrait réfléchir de manière plus développée sur le fait que les émotions puissent être représentées de manière aussi simple.

## 4.5 Pertinence de notre modèle

### 4.5.1 Notre modèle, une représentation d'une conscience humaine

#### Le modèle et la conscience

Ce modèle répond à toutes les définitions de la conscience : Il montre la représentation de la connaissance, la perception et les émotions d'un être vivant conscient, en essayant de se baser sur les travaux de ceux qui ont essayé de la comprendre. Nous pensons qu'il devrait permettre de faire naître l'évolution d'une conscience morale lorsque certains concepts sont rattachés à des notions de bien et de mal en fonction du vécu. Les relations entre les concepts pouvant évoluer, ces notions peuvent évoluer au cours de la vie d'une personne. Toutefois, l'ordre moral étant quelque chose de souvent répété, les notions se cristallisent assez vite lorsque une personne reste dans le même environnement. Enfin, les pensées étant des concepts, il peut les juger pendant leur élaboration, et réfléchir à la pertinence du raisonnement courant.

#### Conscience animale et conscience humaine

Plusieurs personnes ne sont intéressés au problème de conscience animale. Notre modèle est sensé représenter une conscience humaine dans sa globalité, en tenant compte de son évolution. Toutefois, nous ne nions pas l'existence d'une conscience animale. Certains mammifères évolués sont en effet capables d'utiliser des concepts sémantiques, ce qui nécessite une mémoire sémantique et donc un préconscient. Il est en plus probable que certains animaux aient une mémoire épisodique, même si cela reste plus complexe à démontrer. Il reste très difficile de savoir si les animaux arrivent à manier des concepts aussi poussés que ceux des êtres humains, et si ils ont une mémoire de la mémoire. Les solutions pour juger d'un état de conscience d'un être seront discutées dans la partie "Vers une conscience artificielle ?" Mais sans entrer dans le débat de la conscience animale, il faut admettre qu'il n'y a pas forcément que des êtres conscients ou non conscients, mais que la conscience est quelque chose de non absolu, et que chaque être peut être plus ou moins conscient.

### 4.5.2 Le modèle et les phénomènes psychanalytiques

Cette partie n'a pas pour but d'entrer dans les détails des phénomènes étudiés en psychologie et psychanalyse, mais juste de montrer que ce type de phénomènes peuvent s'expliquer sur le modèle aussi bien que sur l'être humain.

#### Les rêves

On peut considérer que les morceaux inconscients qui constituent les rêves, sont des morceaux de vécu qui ont été enfouis profondément dans la mémoire épisodique au point que notre conscient est incapable de les retrouver, mais qui continuent à exister dans la mémoire associative ou le bas de la mémoire sémantique sous forme d'association de concepts. Lors d'un rêve, de la mémoire associative remonte à la mémoire sémantique, mais le conscient n'arrive pas à rattacher de la mémoire épisodique au moment du vécu correspondant. Il rattachera donc d'autres situations inventées à partir d'éléments de la mémoire sémantique et d'autres événements de la mémoire épisodique contenant des concepts communs. Le rêve semble donc parler sous forme de paraboles de choses vécues et oubliées, ou d'idées refoulées mais présentes dans notre inconscient ou notre préconscient. Le passage de ces morceaux pourrait s'expliquer par un besoin d'utiliser ces connaissances oubliées, ou d'une réorganisation de la mémoire sémantique.

#### hypnose et névrose

Nous pouvons aussi expliquer les phénomènes d'hypnose et de névrose, qui sont en réalité assez proche. [Che93] Ils seraient dus à la saturation la mémoire primaire, aux dépens de l'analyseur conceptuel relié indirectement à l'entrée, pour petit à petit oublier ce qui vient de la réalité. La mémoire primaire pouvant directement modifier l'état émotionnel, il peut petit à petit tromper le reste du cerveau sur l'état du corps.

#### Le tabou

Certains troubles mentaux s'expliquent par le fait que le surmoi empêche trop les concepts primaires de remonter. Cela se traduit sur notre modèle par des concepts tabous (associés à la notion morale de mal par la société) que l'analyseur conceptuel poussé ne recopie jamais en mémoire primaire car il est relié à un sentiment de honte. Cela nécessite que certains concepts soient chargés de manière à ce que l'analyseur conceptuel syntaxique évite de les faire passer en mémoire vive, ou à ce que ces concepts soient d'abord rejetés par le choix conscient, et que l'analyseur apprenne de ces rejets le fait qu'il faille arrêter de faire passer ces concepts.

### 4.5.3 Remarques diverses sur le modèle

#### Boot Strap et environnement

Pour que la conscience soit amenée à émerger et à évoluer, il faut que l'environnement dans lequel on la plonge lui soit hostile. Cela la forcerait par nature à vouloir s'adapter à l'environnement ou à l'adapter à lui. Cela lui nécessiterait d'agir, de construire de nouveaux concepts, et donc d'évoluer. De plus, la conscience évoluera plus vite si l'environnement lui envoie des signaux sur plusieurs percepts, de façon à ce qu'elle puisse les mettre en relation. Ainsi, une conscience qui scannerait le web n'aurait qu'un seul type d'entrée et aurait beaucoup de mal à se construire des concepts sémantiques.

#### Les concepts important de la mémoire sémantique

Penser est un acte, et penser est un concept. On peut donc mettre le concept de penser en mémoire primaire, et réfléchir dans le module choix à l'acte de penser. A partir de là, on peut passer en mémoire épisodique une pensée. La méta mnèse permet donc d'arriver à reconstituer le cheminement de la pensée, et on arrive à un niveau meta sur la capacité à juger notre pensée. D'autres concepts sont important, et appuient l'idée que celui qui les possède est conscient : Le concept du moi. Il est probablement nécessaire au concept de pensée. Selon son état d'évolution, il permet de se savoir comme étant un corps dans un environnement. A un plus haut niveau, il peut utiliser les concepts d'identité et permet de se reconnaître comme l'auteur de ses pensées et de ses actes envers son environnement. A ce niveau là, il permet à un individu d'être reconnu

comme responsable de ses actes. Les concept de la société : Pour qu'un individu puisse avoir un comportement social cohérent, il faut qu'il ait des concepts qui lui soient donné par la société. Ces concepts peuvent prendre énormément de poids, et peuvent prendre le pas sur les concepts primaires qui assurent la survie. Ces concepts sont ceux qui vont servir à l'émergence d'un surmoi. Plusieurs maladies mentales peuvent apparaître lorsque le surmoi empêche trop les concepts primaires de remonter. Cela se traduit par des concepts tabous que l'analyseur conceptuel poussé ne recopie jamais en mémoire primaire car il est relié à un sentiment de honte donc à des émotions négatives. Le conscient peut aussi se construire des structures pour préserver son équilibre. Le concept de temporalité est aussi important. il est nécessaire à la méta mnèse. Avec en plus un concept poussé de soit-même, l'individu prend conscience de sa mort. Les concepts de bien et de mal sont nécessaire à l'arrivée d'une conscience morale. D'autre concepts comme le concept de concept, servent à élaborer des réflexions poussées.

### Les limites du modèle

Plusieurs parties du modèle sont encore trop peu définies et manquent de précision. Le "choix conscient" reste assez flou au niveau de la façon dont il pourrait être implémenté. La "somme", module sensé faire la synthèse des sorties des autre modules, devrait pouvoir remonter des informations aux différentes parties du cerveau ou leur demander des informations. D'une manière générale, ce module demanderait une réflexion profonde sur son fonctionnement. Il pourrait entre autre être responsable de la remontée des concepts de plus bas niveau vers la partie consciente du cerveau. La partie sur les émotions reste aussi assez peu satisfaisante. Toutefois, c'est peu être la partie de nous même la plus dur à introspecter, et il sera sans doute difficile de proposer une vision beaucoup plus juste d'implémentation d'émotions dans une structure de conscience artificielle.

#### 4.5.4 Vers une conscience artificielle ?

Cette section va s'intéresser à la question suivante : est ce que le modèle précédent peut aboutir à une véritable conscience ou est-ce peine perdue ? Ce que l'on appelle véritable conscience est une conscience humaine en opposition à une simulation de conscience.

#### La boîte chinoise

Le chercheur Marc Jeannerod a exprimé le fait que quel que soit le comportement observé, qu'il ait nécessité une conscience ou pas, on pouvait le reproduire à partir d'un processus non conscient. Il pense donc que l'on pourra "bluffer" un humain avec un processus qui ne comprend pas ce qu'il fait. L'expérience de l'esprit de la chambre chinoise [Sea85] va dans ce sens. Cette expérience consiste à mettre un français dans une chambre avec un écran sur lequel s'affichent des caractères chinois. Il dispose d'un manuel lui indiquant ce qu'il doit taper sur son clavier lorsque telle suite de caractères s'affiche. Le chinois qui envoie les caractères au français aura l'impression de communiquer avec quelqu'un, pour autant le français ne comprend absolument rien de ce qu'il écrit. On a donc créé une illusion de dialogue. Transposé à l'informatique, un ordinateur pourrait avoir un algorithme qui crée une parfaite illusion de conscience mais pour qui tout cela n'est qu'une suite de données binaires.

#### Naissance d'une véritable conscience

On pourrait penser que, finalement, notre cerveau fonctionne de manière similaire à un ordinateur et manipule des données qui n'ont aucun sens pour lui. Le cerveau utilise des courants électriques qui peuvent être comparés à ceux d'un processeur. Pour le cerveau, manipuler ces courants électriques n'aurait pas plus de sens que pour un ordinateur. Au plus bas niveau, chacun des signaux électriques est interprété comme un simple signal et rien de plus. Cependant, ce qui fait la conscience humaine serait la somme de toutes ces interactions électriques, de tous ces changements d'états. Gestalt a avancé que "le tout est supérieur à la somme des parties", ce qui pourrait parfaitement s'appliquer dans le cas d'une conscience. Cela pourrait amener à penser que, si fondamentalement le processeur ne comprend pas ce qu'il manipule, à une échelle plus globale il y a une logique de haut niveau et finalement il peut y avoir une sorte de conscience. Dans un programme classique, on ne fait que repousser la limite du moment où l'humain s'apercevra qu'il est en face d'une machine. Mais le fait que l'on arrive à créer un programme capable de se comporter comme un humain, sans qu'il ne soit jamais possible de différencier un véritable Humain de la machine,

peut induire que ce programme ne crée plus une illusion mais a atteint un niveau de conscience et que même si aucun des petits processus qui le composent ne comprennent le sens de ce qu'il fait, globalement il est conscient.

### **La Conscience, une propriété de la matière**

Certains scientifiques pensent que la conscience serait une propriété de la matière, et qu'elle ne pourrait être émulée à travers un logiciel. D'autres pensent même que la conscience ne serait pas matérielle, mais serait "autre chose" qui influencerait la matière. Beaucoup de chercheurs, au moment de la découverte de la physique quantique, ont essayé d'adapter les nouvelles connaissances qu'ils ont eu sur la matière aux croyances qu'ils avaient sur la conscience. C'est le cas par exemple de John Carew Eccles auteur de la célèbre citation : « Brain is a machine that ghost can operate » [ER84]. Il pense que notre esprit est non matériel, et est capable d'influencer notre corps en influant sur les phénomènes quantiques présents dans notre cerveau. Eugene Wigner va plus loin en avançant l'idée que notre conscience est la seule chose capable de donner un état fixe aux particules dans un état superposé [Pen05]. Ainsi, il devient impossible de répliquer une conscience ayant des propriétés qui ne pourraient pas être logiciel. Notre conscience artificielle, si elle était implémentée resterait logicielle et ne pourrait avoir les propriétés que certaines personnes prêtent à la conscience humaine. Pour aller plus loin il faudrait d'abord comprendre ces interactions puis les simuler, ce qui n'est pas à l'ordre du jour.

### **Synthèse**

Nous considérerons qu'à partir du moment où un programme se comporterait comme un humain, nous pourrions l'utiliser dans toutes les tâches où nous avons besoin d'une conscience artificielle, et notre objectif serait atteint. Le fait que ce programme puisse par la suite être considéré comme conscient ou non par la communauté de scientifiques et de philosophes ne nous semble pas un critère important dans la réussite de cette tâche.



## Chapitre 5

# L'utilité d'une conscience artificielle

On l'a vu, la question de la faisabilité d'une implémentation de conscience reste ouverte et suscite toujours des débats passionnés. Mais pourquoi dépense-t-on autant d'énergie dans ce but ? Quels sont les intérêts d'une conscience artificielle pour la recherche et les applications concrètes qui en découlent ? La réponse est vaste, et touche de nombreux domaines d'application. On tentera ici de donner quelques pistes de réponse, en cernant quelques champs dans lesquels une conscience artificielle trouverait une utilité immédiate et manifeste.

### 5.1 En IA “classique”

Un logiciel doté de sa propre conscience pourrait se passer de l'interaction humaine tout en restant d'un niveau plus élevé qu'un simple outil algorithmique. Capable de décider ses propres objectifs, d'évaluer ses choix et d'inférer de nouvelles règles pour les atteindre, il permettrait soit de mieux interagir avec l'Homme en s'adaptant mieux à son propre comportement et à ses attentes, soit de s'en passer complètement pour gagner du temps dans certains problèmes. Par exemple, les logiciels de création artistique proposent à l'utilisateur humain des outils pour concevoir des œuvres (de musique ou de dessin, par exemple). Ils n'ont aucune conscience de ce qu'ils font : ils proposent des outils et appliquent des algorithmes, mais c'est l'utilisateur qui crée l'œuvre, sans autre guide que sa propre réflexion. On peut imaginer, avec une conscience dans l'ordinateur, un logiciel non seulement capable de réellement assister le créateur en le guidant dans sa création, mais également capable de proposer et concevoir seul ses propres œuvres, issues de sa propre réflexion. L'interaction homme-machine est un enjeu clé de la conception des interfaces modernes. L'arrivée d'une conscience artificielle chamboulerait complètement ce domaine de recherche, en mettant à égal l'homme et la machine, devenus réellement partenaires dans la résolution d'un problème.

### 5.2 En paradigme agent

Dans un système multi-agents, si tous les agents (humains ou machines) sont dotés d'une conscience propre, ils peuvent non seulement répondre à leurs propres désirs et craintes, mais aussi se représenter les désirs et craintes d'autres agents. Ils deviennent capables de prévoir, d'anticiper les réactions de l'environnement et des autres agents, et de planifier leurs actions en conséquence. Contrairement aux systèmes actuels, les plans issus d'un environnement où tous les agents sont conscients ne sont pas nécessairement guidés par un algorithme : ils peuvent s'appuyer sur les émotions, les désirs de chaque agent, ou même parfois ne pas être rationnels. Rien ne permet d'affirmer qu'ils seraient plus efficaces qu'un plan purement algorithmique, mais en tout cas, ils sont différents car imprévisibles, comme le serait un système d'agents uniquement humains. En outre, la mise en œuvre d'un plan peut aboutir à des organisations en société des agents, chacun prenant en compte les autres non seulement d'après leurs signaux/messages, mais aussi en fonction des concepts qui leur sont associés et de ce qu'on sait de chacun d'eux. Cela devrait permettre un gain d'efficacité et, encore une fois, de résoudre des problèmes qui étaient impossibles à traiter auparavant.

### 5.3 En robotique

Un robot ayant conscience de lui-même et de son environnement et qui peut en déduire de nouvelles règles est théoriquement capable de s'adapter à tout type d'environnement sans qu'il y ait besoin de le reprogrammer. Il peut apprendre quels sont les éléments de l'environnement qui seront source de satisfaction ou de danger pour lui. Mieux : il peut essayer de le deviner et tenter de s'adapter à l'inconnu, en raisonnant sur les concepts qu'il connaît déjà (par analogie, par exemple). Il gagne donc considérablement en autonomie, ce qui augmente le champ de ses possibilités.

### 5.4 Au-delà de la recherche

Le jour où nous pourrions coder une forme de conscience à partir de composants logiciels et électroniques, il deviendra possible de “copier de l'intelligence”. Il suffira de multiplier le nombre de machines à partir d'une seule conscience pour obtenir un réseau d'agents dotés d'une IA forte. La portée des problèmes qu'il deviendrait possible de résoudre s'élargirait sans doute considérablement. Un autre point intéressant est celui de l'évolution d'une conscience artificielle. Comme nous l'avons défini, la conscience est la capacité à évaluer son propre raisonnement et de le modifier en vue de l'améliorer. Une conscience artificielle devrait donc être dotée de cette faculté. Ainsi, à partir d'une seule conscience, il deviendrait possible d'obtenir une intelligence capable de s'améliorer elle-même, et ce de façon récursive (la “nouvelle” pensée étant à son tour évaluée, puis améliorée, en améliorant éventuellement le processus d'évaluation lui-même). On peut même imaginer une conscience capable de créer elle-même une nouvelle conscience sur des bases entièrement nouvelles, issues de sa propre réflexion ! Combinée à la puissance et à la duplication des machines, cette possibilité semble repousser les frontières du possible vers des limites impossibles à imaginer.

## Chapitre 6

### conclusion

La conscience artificielle est un domaine encore peu exploré, qui soulève des questions philosophiques et pose des problèmes informatique d'implémentation, mais sa création pourrait révolutionner le monde de l'informatique. malgré toutes les difficultés soulevés au cours de ce mémoire, nous pensons qu'il est possible d'implémenter une conscience artificielle, et que c'est quelque chose de nécessaire pour passer dans une nouvelle ère informatique.



# Bibliographie

- [AS68] R.C. Atkinson and R.M. Shiffrin. Human memory : A proposed system and its control processes. In *The psychology of learning and motivation (Volume 2)*. Academic Press, 1968.
- [Car99] Alain Cardon. *Conscience artificielle et systèmes adaptatifs*. Eyrolles, 1999.
- [Che93] Léon Chertok. *Hypnose et suggestion*. Presses universitaires de France, 1993.
- [CQ95] Allan M. Collins and M. Ross Quillian. Retrieval time from semantic memory. In *Computation & intelligence*. American Association for Artificial Intelligence, 1995.
- [ER84] Sir John Eccles and Daniel N. Robinson. *The Wonder of Being Human : Our Brain and Our Mind*. Free Press, 1984.
- [Fre23] Sigmund Freud. *Le moi et le ça*. Payot, 1923.
- [Kur06] Ray Kurzweil. *The Singularity Is Near : When Humans Transcend Biology*. Penguin (Non-Classics), 2006.
- [Lab76] Henry Laborit. *Éloge de la fuite*. Editions Robert Laffont, 1976.
- [Lec97] Christophe Lecerf. Une leçon de piano. *TRAVAUX ET DOCUMENTS*, 1997.
- [PAN01] Chantal COMBE PANGAUD. *CONTRIBUTION A L'ETUDE DE LA COMPLEXITE DU SYSTEME MNESIQUE HUMAIN*. PhD thesis, UNIVERSITE LUMIERE LYON 2, 2001.
- [Pen05] Roger Penrose. *The Road to Reality : A Complete Guide to the Laws of the Universe*. Alfred A. Knopf, 2005.
- [Sab90] Gérard Sabah. Caramel : a flexible model for interaction between the cognitive processes underlying natural language understanding. In *Proceedings of the 13th conference on Computational linguistics - Volume 3, COLING '90*. Association for Computational Linguistics, 1990.
- [Sab93] Gérard Sabah. Vers une conscience artificielle. In *modèles et concepts pour la science cognitive*. Presse Universitaire de Grenoble, 1993.
- [Sea85] John R. Searle. Minds, brains, and programs. In *Mind design*. MIT Press, 1985.
- [Sow84] J. F. Sowa. *Conceptual structures : information processing in mind and machine*. Addison-Wesley Longman Publishing Co., Inc., 1984.
- [SVV<sup>+</sup>97] Gérard Sabah, Jean Vivier, Anne Vilnat, Jean-Marie Pierrel, Laurent Romary, and Anne Nicolle. *Machine, langage et Dialogue*. Figure de l'interaction, 1997.
- [Tul72] Eden Tulving. *Organization of Memory*. Academic Press, 1972.
- [VNP02] Rémi Versace, Brigitte Nevers, and Catherine Padovan. *La mémoire dans tous ses états*. Solal, 2002.