



**HAL**  
open science

## Les concepts ne sont pas une espece naturelle

Edouard Machery

► **To cite this version:**

Edouard Machery. Les concepts ne sont pas une espece naturelle. domain\_scco.cogpsy. Université Paris-Sorbonne - Paris IV, 2004. Français. NNT : . ijn\_00000514

**HAL Id: ijn\_00000514**

**[https://theses.hal.science/ijn\\_00000514](https://theses.hal.science/ijn_00000514)**

Submitted on 29 Jun 2004

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS IV - SORBONNE

ÉCOLE DOCTORALE CONCEPTS ET LANGAGES

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS IV

Discipline : Philosophie

présentée et soutenue publiquement par

Edouard MACHERY

Le 28 juin 2004

**LES CONCEPTS NE SONT PAS UNE ESPÈCE  
NATURELLE**

LA NOTION DE CONCEPT EN PSYCHOLOGIE COGNITIVE

Directeur de thèse : M. Daniel ANDLER (professeur, Paris IV)

**Jury**

M. Pascal Engel (professeur, Université Paris IV)

M. James Hampton (professeur, City University London)

M. Dan Sperber (directeur de recherche, Institut Jean-Nicod, CNRS)

## Remerciements

---

Comme de nombreux doctorants, dans ma première année de thèse (2000-2001), j'ai parcouru, plus ou moins cursivement, plusieurs thèses afin de me faire une idée un peu plus précise du travail que l'on attendait de moi. Et, comme beaucoup, sans doute, je ne pus m'empêcher de trouver un peu ridicules, parce que fort conventionnels, les remerciements qui, invariablement, précédaient ces travaux. Toutefois, maintenant que c'est à moi de présenter le résultat de quatre années de travail, je comprends que ce jugement hâtif était foncièrement erroné. Certes, les remerciements ont quelque chose de conventionnel. Mais ils ne sont en rien ridicules. Car on témoigne ainsi de sa reconnaissance envers tous ceux sans qui son propre travail n'aurait jamais pu aboutir.

Remercions tout d'abord les nombreux anonymes. De nombreux aspects de ce travail ont en effet été présentés lors de colloques et de conférences. Les commentaires, questions et critiques qui m'ont été adressés m'ont permis d'améliorer ce travail. Je remercie tous ceux qui ont eu la patience de m'écouter et la gentillesse de me suggérer des améliorations.

J'ai de très nombreuses dettes intellectuelles. Je dois mon intérêt pour la psychologie cognitive à mon directeur de thèse, Daniel Andler. Je lui suis fort reconnaissant d'avoir été si patient et d'avoir approuvé aussi bien mon séjour outre-Atlantique que le temps consacré à mes autres intérêts philosophiques. Je lui suis tout particulièrement redevable du temps qu'il m'a consacré lors de la phase finale de ce travail. Sa lecture attentive et sans concession m'a permis d'approfondir certaines questions, d'écartier bien des ambiguïtés et d'éviter de nombreuses fautes. Dès la création de l'Institut Jean-Nicod, j'ai eu la chance d'être associé de près aux travaux qui y étaient menés, en premier lieu par François Récanati et Jérôme Dokic au sein du séminaire Index. Philippe Schlenker, pendant son année d'enseignement à l'École Normale Supérieure, a orienté mes intérêts alors que je découvrais à peine la philosophie analytique. Même si ce travail doit finalement peu à la linguistique et beaucoup à la psychologie, je lui suis redevable de m'avoir fait découvrir la philosophie du langage, la sémantique et les rudiments de la syntaxe. Je dois finalement beaucoup à Steve Stich.

Grâce à sa générosité, j'ai pu passer plus d'un an et demi à Rutgers. C'est peu dire que j'y ai modifié de fond en comble ma manière de faire de la philosophie. C'est peu dire que mes intérêts philosophiques ont été intégralement chamboulés. Quoique ce travail ne corresponde pas tout à fait au type de travaux recommandé par Steve Stich à ses étudiants, il aurait une toute autre allure sans son influence. Les discussions avec les étudiants et les professeurs du département de philosophie et du centre de sciences cognitives de Rutgers ont constamment enrichi ma réflexion. Enfin, mon séjour à l'Institut Max-Planck pour le Développement Humain, à Berlin, m'a permis, entre autres choses, de finir ce travail. Je remercie le DAAD de m'avoir accordé une bourse de recherche à cette fin.

Ma plus grande dette va finalement à mes parents. C'est le moment idéal pour leur témoigner toute ma gratitude. *Merci*. Je souhaite aussi les remercier, ainsi que ma grand-mère, de leur aide pendant la phase finale de ce travail. J'ai la chance d'avoir une famille unie et disposée à m'aider dans les moments les plus difficiles. Je souhaite aussi remercier Nastya, en compagnie de qui j'ai passé mes trois dernières années de thèse et sans qui la vie serait beaucoup moins plaisante.

# Sommaire

---

<b>Introduction</b>	<b>5</b>
<b>Chapitre 1 Qu'est-ce qu'une théorie des concepts ?</b>	<b>17</b>
<b>Chapitre 2 Les concepts ne forment pas une espèce naturelle</b>	<b>59</b>
<b>Chapitre 3 Trois types de concepts</b>	<b>83</b>
<b>Chapitre 4 L'hétérogénéité conceptuelle</b>	<b>109</b>
<b>Chapitre 5 L'hétérogénéité de la catégorisation</b>	<b>143</b>
<b>Chapitre 6 La composition des concepts</b>	<b>196</b>
<b>Chapitre 7 Est-ce que l'eau est H<sub>2</sub>O ?</b>	<b>226</b>
<b>Chapitre 8 Contre les théories composites des concepts</b>	<b>255</b>
<b>Conclusion</b>	<b>280</b>
<b>Bibliographie</b>	<b>307</b>
<b>Abréviations</b>	<b>333</b>
<b>Index</b>	<b>334</b>
<b>Tables des matières</b>	<b>339</b>

## Introduction

---

Livres, articles, conférences, tables rondes, séminaires... La philosophie des concepts est l'un des champs les plus dynamiques de la philosophie analytique contemporaine. Pourtant, les controverses qui l'agitent tournent souvent autour d'un petit nombre de problèmes. En particulier, deux problèmes ont été au cœur des débats philosophiques des quinze dernières années, l'individuation des concepts et la psychosémantique. On peut caractériser ces problèmes comme suit.

- *l'individuation des concepts* : les philosophes des concepts ont cherché à répondre à la question « Qu'est-ce qui distingue un concept de x d'un concept de y ? », par exemple, un concept de chien d'un concept de chat ou un concept de triangle d'un concept de trilatéral ;
- *l'explication des propriétés sémantiques des concepts* : les philosophes des concepts ont aussi cherché à répondre à la question « Comment un concept réfère-t-il à ce qu'il réfère ? », par exemple, comment CHIEN réfère-t-il à la classe des chiens ?<sup>1</sup>

Quoique ces questions aient été au cœur de la philosophie des concepts, j'ai choisi de les négliger dans ce travail. De nombreuses raisons motivent cette orientation.

Il est extrêmement délicat d'apporter une réponse nouvelle à ces questions, dans la mesure où la plupart des positions concevables ont été examinées. Au mieux, on peut espérer développer les détails d'une position déjà présente dans le champ philosophique ou avancer de nouvelles objections contre les positions déjà formulées. Plutôt que de m'inscrire dans un espace dont la cartographie est déjà bien connue, j'ai préféré examiner un nouveau problème.

La philosophie des concepts fait traditionnellement partie de la philosophie de l'esprit. Or cette dernière discipline, qui était à l'origine fortement inspirée par la

---

<sup>1</sup> Dans ce qui suit, les noms des concepts sont écrits en lettres capitales. Les noms des propriétés sont écrits en italiques. Les noms de propositions sont écrits entre crochets.

psychologie et les sciences cognitives en général, s'est progressivement éloignée de ces disciplines scientifiques. Des préoccupations métaphysiques et épistémologiques sont aujourd'hui monnaie courante en philosophie de l'esprit. La philosophie des concepts ne fait pas exception. Elle est souvent liée à la théorie de la justification (Peacocke 1992, 1999, McDowell 1994, Brandom 2001), tandis que la différence entre la psychosémantique et la métaphysique est parfois ténue. Mon goût philosophique me porte au contraire vers la philosophie des sciences. A mes yeux, l'une des tâches de la philosophie des sciences consiste à examiner la pertinence de notions scientifiques. Une telle tâche ne peut être accomplie *a priori*, dans la mesure où la pertinence des notions scientifiques est fonction des phénomènes empiriques qu'elles sont supposées expliquer. La philosophie de la psychologie est une sous-discipline de la philosophie des sciences : elle s'intéresse aux notions, méthodes, théories etc. qui structurent la psychologie. J'ai préféré étudier les concepts dans le cadre de philosophie de la psychologie plutôt que dans le cadre de la philosophie de l'esprit. Je prends donc pour point de départ la notion de concept utilisée en psychologie cognitive. Et je cherche à montrer que la recherche sur les concepts menée en psychologie cognitive soulève de nombreuses questions qui, malgré leur importance philosophique, ont été largement négligées par les philosophes.

Finalement, la méthodologie utilisée pour résoudre les problèmes que pose la philosophie des concepts me semble critiquable. Les arguments avancés dans les domaines de la psychosémantique et de la théorie de la justification reposent essentiellement sur des intuitions (par exemple, Burge 1986, Davies 2000). Lorsque l'on traite ces intuitions de manière naturaliste, comme des phénomènes cognitifs, la portée de ces arguments devient problématique (Stich 1990, 1996, Weinberg et al. 2001, Stich et Weinberg 2001, Machery et al. 2004). Parmi les nombreux problèmes que pose le recours aux intuitions, notons en particulier le problème suivant. Certaines intuitions épistémologiques et sémantiques varient de manière prévisible d'une zone culturelle à l'autre. Ce fait est difficilement compatible avec le statut qui est accordé à ces intuitions en philosophie. Pourquoi en effet devrait-on accorder un statut particulier aux intuitions d'un groupe culturel par rapport aux intuitions d'un autre groupe culturel ? Au contraire, la discussion des problèmes posés par la notion psychologique de concept ne fait pas appel à nos intuitions.

Le lecteur ne doit donc pas être surpris si les problèmes habituellement traités en philosophie de l'esprit ne sont qu'esquissés dans ce travail – souvent pour être aussitôt mis à l'écart. Je l'invite même à faire abstraction de ces problèmes. L'un des aspects les plus ambitieux de ce projet consiste en effet à déplacer le centre de gravité du questionnement philosophique sur les concepts – ou, pour utiliser une autre métaphore, à redessiner un peu la carte des problèmes philosophiques que pose la notion de concept. De fait, l'activité philosophique consiste autant à définir les problèmes qu'à y répondre.

Ce travail s'inscrit donc dans le cadre de la philosophie de la psychologie, et, plus généralement de la philosophie des sciences. Il existe plusieurs méthodes pour mener à bien un projet philosophique à propos d'une science particulière, par exemple en philosophie de la biologie ou en philosophie de la psychologie. Le philosophe peut se placer à l'extérieur du champ disciplinaire qu'il étudie, dans le cas de ce travail, à l'extérieur de la psychologie cognitive (voir, par exemple, Descombes 1995 sur la psychologie cognitive). La perspective est souvent critique : le philosophe cherche à formuler les hypothèses fondamentales qui structurent un champ disciplinaire et à en examiner la validité. Le philosophe peut au contraire se placer à l'intérieur du champ disciplinaire qu'il étudie (voir, par exemple, Griffith 1997 et Sober et Wilson 1998 en philosophie de la psychologie et de la biologie).<sup>2</sup> Il admet de nombreuses thèses et notions de ce champ. Souvent, une telle acceptation est purement pragmatique. Le philosophe admet ces notions pour pouvoir mener à bien un argument particulier, tout en étant disposé à reconnaître qu'elles mériteraient un examen critique. Dans la mesure où le philosophe accepte de nombreuses notions et principes théoriques, son travail se distingue alors parfois difficilement d'une élaboration théorique au sein de la discipline concernée (par exemple, voir Sober et Wilson 1998 sur l'évolution et la psychologie de l'altruisme).

Les deux approches sont sans aucun doute légitimes et complémentaires. Toutefois, on ne peut les mener de front. En conséquence, dans ce travail, j'adopte pleinement la seconde perspective. J'admets, sans les discuter, de nombreuses notions et thèses de la psychologie des concepts et des fonctions cognitives supérieures, thèses et

---

<sup>2</sup> Ce sont bien sûr des idéaux types.

notions que je critiquerais sans doute si j'avais adopté une autre approche. Certaines de ces notions jouent un rôle central dans l'argument développé dans ce travail. C'est le cas par exemple de la notion d'encodage, de la notion de mémoire à long terme, de la notion d'extraction ou de la notion de catégorisation. Afin de pouvoir mener à bien l'argument central de ce travail, je n'examine pas la validité de ces notions. Le lecteur ne doit donc pas s'attendre à un examen critique de toutes les notions et thèses mentionnées dans ce travail.

Par certains aspects, ce travail s'apparente donc à de la théorisation psychologique. Je suppose qu'un psychologue, intéressé par des questions théoriques, aurait pu écrire de nombreux passages de ce travail. Cela s'accorde bien avec la perspective naturaliste qui est la mienne : la distinction entre la philosophie des sciences et les parties théoriques des sciences est souvent purement conventionnelle. La philosophie des sciences est donc souvent de même nature que la théorisation scientifique. En outre, on ne devra pas s'étonner si une large part de ce travail est consacrée à la recension, à l'examen et à la mise en perspective de données empiriques. Souvent, on ne peut mener à bien un travail en philosophie des sciences, en particulier à propos d'une science particulière, sans prendre en considération de nombreuses données empiriques.

De toute évidence, le terme « concept » est aujourd'hui utilisé dans de nombreuses disciplines, en psychologie et en philosophie bien sûr, mais aussi en histoire, par exemple en histoire des sciences (le concept de phlogistique, le concept de masse etc.) ou encore en anthropologie. Ces usages du terme « concept » sont de toute évidence liés les uns aux autres : dans chaque cas, on insiste sur la manière dont un individu ou un groupe d'individus pense à un objet, à une classe d'objets etc. Cependant, dans de nombreuses disciplines, la notion a peu de contenu théorique : le terme « concept » est simplement équivalent à « manière de penser à ». C'est sans doute le cas de l'histoire ou de l'anthropologie. Dans ce qui suit, j'adhère exclusivement à l'usage *psychologique* du terme « concept ». Comme nous allons le voir ci-dessous, les concepts en psychologie cognitive sont des ensembles de connaissances à propos de classes d'objets qui sont stockés en mémoire à long terme (LTM) et qui sont utilisés par défaut dans nos processus

cognitifs supérieurs. La notion de concept doit donc être vue comme *une notion théorique*, introduite par une discipline scientifique, à savoir la psychologie des fonctions cognitives supérieures, à des fins explicatives (voir aussi Millikan 1998). En conséquence, nos intuitions préthéoriques ou philosophiques sur la notion de concept n'ont guère de pertinence – à supposer que nous ayons des intuitions préthéoriques à ce sujet. En particulier, nos intuitions sur ce qui peut légitimement être tenu pour un concept n'ont pas plus de pertinence pour étudier la notion de concept en psychologie cognitive que nos intuitions sur ce qu'est la masse pour étudier la notion de masse dans la théorie de la relativité. J'invite donc le lecteur à mettre de côté ces intuitions préthéoriques ou philosophiques à propos des concepts.

S'il est vrai que le terme « concept » est traité dans ce travail comme un terme théorique introduit par de nombreuses théories scientifiques en psychologie cognitive, on ne peut toutefois ignorer complètement son origine philosophique. Même s'il ne peut être question ici de faire un historique détaillé de cette notion, il importe de résumer son histoire.<sup>3</sup> L'intérêt pour les représentations mentales des individus est aussi ancien que la philosophie elle-même. Platon et Aristote s'interrogent en effet sur la nature de nos pensées et distinguent différents types de pensées (voir, par exemple, *La République* de Platon et le *De Anima* d'Aristote). Cette interrogation psychologique est inséparable d'une interrogation de nature épistémologique. Autrement dit, les pensées sont étudiées non seulement pour leurs propriétés psychologiques – comment nous les formons, comment nous les utilisons etc., mais aussi pour leur statut épistémologique – quel type de pensée conduit à la connaissance ? Ce double questionnement, psychologique et épistémologique, traverse toute l'histoire de la philosophie et se retrouve encore dans la philosophie contemporaine des concepts (Peacocke 1992, McDowell 1994). A la suite des philosophes du Moyen Age (voir, par exemple, de Libera 1997), les philosophes modernes des dix-septième et dix-huitième siècles étudient aussi la nature de nos pensées – le terme « idée » étant un précurseur du terme plus moderne « concept ». Ils posent et revisitent de nombreux problèmes, dont certains sont encore discutés de nos jours, aussi bien en philosophie qu'en psychologie. Les concepts sont-ils innés ? Sont-ils dérivés de nos perceptions ? Les concepts sont-ils semblables à nos perceptions ? Si oui, en quel

---

<sup>3</sup> Ce qui suit est une description subjective, mais, je l'espère, fidèle de l'histoire du terme « concept ».

sens ? Si non, quelle est la différence ? Les concepts encodent-ils une information générale ? Qu'est-ce que l'abstraction ? Quelle est la nature des concepts abstraits ? Descartes, Locke, Berkeley, Leibniz et Hume examinent ces questions et établissent en partie l'agenda de la philosophie et de la psychologie des concepts des siècles suivants. Notons que, comme Platon et Aristote, leur réflexion est aussi bien guidée par un intérêt psychologique que par des préoccupations épistémologiques sur la nature et la possibilité de la connaissance.

Le terme « concept » lui-même semble avoir été introduit par les philosophes de langue allemande (« Begriff »).<sup>4</sup> En particulier, le terme « concept » remplace chez Kant le terme « idée » utilisé par les philosophes des dix-septième et dix-huitième siècles pour dénoter les représentations de classes d'objets impliquées dans les raisonnements. De nouveau, la perspective épistémologique est inséparable de la perspective psychologique : Kant s'intéresse surtout aux concepts afin de pouvoir expliquer la nature et la possibilité de la connaissance (voir *La Critique de la Raison Pure*). Ce terme est bien sûr repris par l'idéalisme post-kantien, en particulier par Hegel : l'aspect épistémologique de la notion est particulièrement développé par ce dernier, puisque le concept est supposé être la forme que prend l'esprit lorsqu'il atteint la pleine connaissance de la réalité (voir *La Phénoménologie de l'Esprit*).

Dès la création de la psychologie expérimentale, dans la seconde moitié du dix-neuvième siècle, les psychologues cherchent à déterminer la nature des concepts (Wundt 1874, Thorndike 1913, voir la recension des premiers pas de la psychologie des concepts dans Fisher 1916, 2-30). Ils héritent sans doute le terme « concept » de Kant et des idéalistes allemands (« Begriff »). Comme Kant, et comme les philosophes pré-kantiens avec le terme « idée », ces psychologues désignent par « concept » les représentations mentales de classes d'objets. Ils cherchent à déterminer la nature de ces représentations et la manière dont elles sont formées à partir de nos perceptions (Fisher 1916 résume l'ensemble des recherches sur les concepts et l'abstraction qui ont été menées depuis l'invention de la psychologie expérimentale).<sup>5</sup> Ils revisitent les questions posées par les philosophes des dix-septième et dix-huitième siècles ainsi que par Kant. Néanmoins,

---

<sup>4</sup> Ce n'est là qu'une conjecture.

<sup>5</sup> Voir aussi Machery, E., « Concept theories : Old ideas in disguise », manuscrit.

contrairement aux philosophes, l'interrogation épistémologique sur la nature et la possibilité de la connaissance n'est pas au cœur de leur interrogation.<sup>6</sup> Leur questionnement est surtout de nature psychologique. Malgré le rejet des notions mentalistes par les béhavioristes, le terme « concept » ne disparaît pas de la psychologie (par exemple, Smoke 1932, voir aussi Murphy 2002, 11-15). La notion de concept y est redéfinie comme une réponse suscitée par les membres d'une classe d'objets. Cependant, avec le retour en grâce des notions mentalistes dans les années cinquante, l'étude de la nature et de l'acquisition des représentations mentales de classes d'objets attire l'attention de nombreux chercheurs (voir, par exemple, Bruner et al. 1956, voir aussi les recensions de Smith et Medin 1981, Medin 1989, Komatsu 1992 et Murphy 2002, chap. 1). L'étude des concepts est alors inséparable de l'étude de leur acquisition. On sait l'importance des années 70 dans la psychologie des concepts et des fonctions cognitives supérieures, en premier lieu de la catégorisation. Dans les années 70, de nouvelles théories des concepts sont développées (voir le chapitre 3) et la psychologie des concepts se renouvelle pour une très grande part (voir Smith et Medin 1981, Medin 1989, Komatsu 1992, Margolis 1994, Hampton 1997a, Laurence et Margolis 1999 et Murphy 2002). Mon travail porte sur les théories qui sont nées depuis les années soixante-dix.

La philosophie n'a de toute évidence pas abandonné la notion de concept. A la fin du dix-neuvième siècle et au début du vingtième siècle, Frege tenta de séparer la problématique psychologique (quelle est la nature des états mentaux qui représentent des classes d'objets ?) de la notion de concept (Frege 1894/1994). Le terme « concept » ne réfère pas chez Frege à nos états mentaux, mais à des entités abstraites qui sont exprimées par les termes généraux, par exemple par « chat ». Ces entités abstraites constituent le sens des expressions qui les expriment et sont appréhendées par l'esprit. Le terme « concept » fut donc associé à une problématique sémantique : doit-on postuler ce type d'entités abstraites afin de caractériser la signification des termes généraux ? C'est en ce sens que le terme « concept » fut utilisé par certains philosophes du vingtième siècle, par

---

<sup>6</sup> Toutefois, on pourrait soutenir qu'elle n'est pas entièrement absente. En effet, dans les paradigmes expérimentaux utilisés par les psychologues de la fin du dix-neuvième siècle et du début du vingtième siècle (Grünbaum 1908, Moore 1910, Fisher 1916, Hull 1920), acquérir un concept consiste à trouver la propriété qui est *réellement* commune à l'ensemble des objets utilisés dans l'expérience. Cette idée est aussi présente dans la manière dont Piaget utilise la notion de concept (Piaget 1953) : selon Piaget, les enfants acquièrent des concepts quand leur pensée satisfait certaines exigences épistémiques.

exemple par Carnap (Carnap 1947). De manière frappante, dans la philosophie contemporaine, les trois problématiques philosophiques évoquées ci-dessus, à savoir les problématiques épistémologique, sémantique et psychologique, sont étroitement associées. Les concepts sont les composants de nos attitudes propositionnelles, croyances, désirs etc., ils constituent le sens des termes généraux et ils ont une signification épistémologique (voir, par exemple, Evans 1982). Au contraire, seule la problématique psychologique est aujourd'hui pertinente en psychologie cognitive. Et c'est elle qui est au cœur de ce travail.

Tournons-nous maintenant résolument vers *la psychologie contemporaine des concepts* et des fonctions cognitives supérieures. Précisons, tout d'abord, ce que désigne, dans ce travail, les expressions « psychologie des concepts » et « psychologie des fonctions cognitives supérieures ». De nombreuses sous-disciplines psychologiques étudient les fonctions cognitives supérieures.<sup>7</sup> Elles se distinguent les unes des autres par leurs méthodes ou par les problèmes qu'elles examinent. Nombre de ces disciplines utilisent régulièrement la notion de concept. C'est le cas de la psychologie expérimentale, qu'elle porte sur le raisonnement, sur notre capacité de résoudre des problèmes, sur la catégorisation ou sur notre capacité de faire des analogies. C'est aussi le cas de la psychologie cognitive développementale (Carey, Spelke, Bloom, Mandler etc.), de l'étude de certaines psychopathologies, comme l'amnésie<sup>8</sup>, la démence sémantique ou les déficits propres à certaines catégories<sup>9</sup> (« category-specific deficits ») et de la neuroimagerie fonctionnelle<sup>10</sup>. Dans ces sous-disciplines, de nombreux psychologues soutiennent explicitement que l'objet de leur travail est de mettre au jour la nature de nos concepts. Le terme « psychologie des concepts » désigne ainsi souvent l'activité des psychologues qui étudient nos fonctions cognitives supérieures en cherchant particulièrement à déterminer la nature des concepts (voir, par exemple, Murphy 2002, 2). Dans le présent travail, je me concentre particulièrement sur la psychologie

---

<sup>7</sup> Sur la notion de fonction cognitive supérieure, voir le chapitre 1, section 1.2.1.

<sup>8</sup> Pour des travaux sur l'amnésie qui sont pertinents pour le présent travail, voir, entre autres, Kolodny 1994, Knowlton et Squire 1993, Squire et Knowlton 1995, Kitchener et Squire 2000 et Smith et Minda 2000b, ainsi que la courte recension de Knowlton 1999.

<sup>9</sup> Voir les recensions de Caramazza et Mahon 2003, Mahon et Caramazza 2003, Moss et Tyler 2003.

<sup>10</sup> Voir, par exemple, Smith et al. 1998, Reber et al. 1998, Smith et Jonides 1999.

expérimentale qui s'est développée depuis les années 70, même si, à l'occasion, je fais de brèves excursions dans la psychologie développementale ou l'étude expérimentale des psychopathologies. Cette restriction résulte autant des limites naturelles d'un travail de thèse, qui ne saurait prétendre à l'exhaustivité, que de la nature même de la psychologie des concepts. En effet, pour une grande part, la psychologie des concepts s'est développée au sein de la psychologie expérimentale.

Comme nous allons le voir en détail, la plupart des psychologues des concepts et des fonctions cognitives supérieures admettent une image particulière de notre cognition supérieure. Nous avons en mémoire des ensembles de connaissances relativement bien définis à propos de classes d'entités qui sont utilisés par défaut dans un grand nombre de tâches cognitives (catégorisation, induction, déduction...). Ce sont nos *concepts*, et la psychologie des concepts cherche à les décrire. L'étude des concepts est par ailleurs souvent vue comme la clef pour comprendre les fonctions cognitives supérieures de l'esprit. Nombreux sont ceux qui supposent que déterminer la nature des ensembles de connaissances sur lesquels reposent nos compétences cognitives supérieures serait un grand pas vers la compréhension de ces compétences.

Dans la mesure où cette image de l'esprit est admise par de nombreux psychologues des concepts, elle ne peut manquer d'influencer les recherches qu'ils mènent, la manière dont ils interprètent leurs résultats empiriques et les débats et controverses qui agitent la psychologie des concepts et la psychologie des fonctions cognitives supérieures. Tout en admettant la plupart des notions et principes des psychologues cognitifs, j'examine de près certains aspects de cette image de l'esprit.

Depuis les années 70, les psychologues ont proposé de nombreuses théories des concepts. Je soutiens que les psychologues des concepts et des facultés cognitives supérieures humaines, par exemple de la catégorisation ou du raisonnement inductif, ont en général supposé que *les concepts forment une espèce naturelle* : autrement dit, ils forment une classe d'entités qui possèdent de manière non accidentelle de nombreuses propriétés communes et donnent en conséquence lieu à de nombreuses généralisations empiriques. La notion de concept est une manière pertinente d'organiser les représentations mentales des humains pour une psychologie scientifique. Bref, les

psychologues croient avoir trouvé avec la notion de concept une manière importante de classer nos états mentaux pour la psychologie scientifique. Nier que les concepts forment une espèce naturelle, c'est au contraire affirmer qu'il n'est pas possible de formuler des généralisations empiriques à leur propos. Si cela était le cas, la classe des concepts n'aurait probablement qu'une importance limitée pour la psychologie cognitive. Les psychologues auraient donc eu tort de mettre la notion de concept au centre de leur travail. D'autres notions seraient requises. C'est précisément là la thèse centrale de ce travail : *les concepts ne forment pas une espèce naturelle*.

Ce travail comporte neuf chapitres. Je décris brièvement ci-dessous leur contenu.

- **Chapitre 1 Qu'est-ce qu'une théorie des concepts ? (prémisse 1 de l'ANE)**

Dans le premier chapitre, j'introduis la notion de concept et je décris l'objet et les méthodes des théories empiriques des concepts. J'insiste sur les différences entre la manière dont les philosophes de l'esprit ont caractérisé les concepts et ce que sont les concepts pour les psychologues cognitifs. Je montre en particulier, contre Jerry Fodor, que les concepts sont postulés pour expliquer les propriétés des processus cognitifs.

- **Chapitre 2 Les concepts ne forment pas une espèce naturelle**

Le deuxième chapitre est consacré à la présentation détaillée de ma position. Après avoir analysé la notion d'espèce naturelle, j'explique pourquoi je rejette l'idée que les concepts forment une espèce naturelle et j'expose ma stratégie argumentative, que j'appelle *l'argument de la nécessité explicative* (ANE). En bref, cet argument repose sur l'idée qu'il faut postuler l'existence de plusieurs types de concepts qui ont peu en commun – ce que j'appelle « *l'hypothèse de l'hétérogénéité* » – si l'on souhaite rendre compte de l'ensemble des propriétés des compétences cognitives supérieures. Finalement, je compare ma position à des thèses plus ou moins semblables avancées par certains psychologues.

- **Chapitre 3 Trois types de concepts**

Dans le troisième chapitre, j'expose les types de concepts dont il faut, à mes yeux, postuler l'existence. Je m'intéresse en particulier aux prototypes, aux exemplaires et aux

théories. Je présente aussi d'autres types de représentations mentales, en particulier les symboles perceptuels de Barsalou et de Prinz.

- **Chapitre 4 L'hétérogénéité conceptuelle (prémisse 3 de l'ANE)**

Dans le quatrième chapitre, je montre que les prototypes, les exemplaires et les théories sont des types de représentations mentales qui ont peu en commun. En bref, je soutiens que ces trois entités théoriques ont des propriétés sémantiques et fonctionnelles différentes. Je critique finalement un argument avancé par Barsalou (1990) selon lequel les prototypes et les exemplaires ne peuvent être distingués expérimentalement.

- **Chapitre 5 L'hétérogénéité de la catégorisation**

Dans le cinquième chapitre, j'examine les travaux empiriques sur la catégorisation. Dans la lignée des théories duelles des processus cognitifs, je propose qu'il existe plusieurs processus cognitifs pour catégoriser. Ces processus font appel à différents types de représentations, en particulier aux prototypes, aux exemplaires et aux théories. Cette hypothèse permet d'expliquer la diversité des propriétés qui caractérisent nos catégorisations. Elle conduit aussi à poser de nouvelles questions sur l'organisation de nos processus cognitifs – questions qui mériteraient d'être examinées de près par les psychologues.

- **Chapitre 6 La composition des concepts**

Le sixième chapitre est consacré à la littérature empirique sur la composition des concepts. Je propose que cette littérature suggère l'existence aussi bien de prototypes que d'exemplaires et de théories. Je décris plusieurs phénomènes empiriques dont l'explication requiert ces différents types de représentations mentales. Je propose aussi un mécanisme cognitif qui peut expliquer la création de concepts complexes.

- **Chapitre 7 Est-ce que l'eau est H<sub>2</sub>O ?**

Dans le septième chapitre, j'étudie l'essentialisme psychologique et les travaux critiques que cette théorie a suscités (Malt, Hampton...). Je montre que la contradiction apparente entre ces travaux peut être résolue si l'on admet l'existence de plusieurs types de

concepts. Je montre que nous avons plusieurs manières de conceptualiser ce qu'est l'eau, c'est-à-dire plusieurs concepts d'eau. L'un d'entre eux est théorique : il inclut la croyance essentialiste que l'eau est essentiellement faite de molécules dont la structure est H<sub>2</sub>O. Un autre semble être un prototype ou un ensemble d'exemplaires d'eaux.

- **Chapitre 8 Contre les théories composites des concepts**

Dans le huitième chapitre, j'examine et critique les théories composites des concepts (Osherson et Smith, Armstrong, Gleitman et Gleitman, Gelman). Ces théories supposent que les concepts sont composés de plusieurs parties qui encodent des connaissances différentes et qui sont utilisées dans des processus cognitifs différents. Ces théories semblent à même d'expliquer la plupart des résultats empiriques qui selon moi, viennent à l'appui de l'hypothèse de l'hétérogénéité. Pour défendre l'hypothèse de l'hétérogénéité, je propose un argument contre ces théories : je montre qu'elles sont ou bien incompatibles avec certaines données empiriques ou bien de simples variantes terminologiques de l'hypothèse de l'hétérogénéité.

- **Conclusion**

Dans le dernier chapitre, je résume les acquis négatifs et positifs de ce travail. Il faut admettre l'existence de plusieurs types de concepts pour expliquer les propriétés des compétences cognitives supérieures, en particulier l'existence de prototypes, d'exemplaires et de théories. Comme ces types de concepts ont très peu en commun, je conclus que les psychologues ont eu tort de supposer que les concepts formaient une espèce naturelle. Plutôt que de prétendre décrire les concepts en général, ils devraient donc se concentrer sur des types particuliers de concepts. En outre, les controverses entre les partisans des différentes théories des concepts n'ont guère lieu d'être, puisque ces théories portent en fait sur des types différents de concepts. Je souligne aussi qu'adopter ma proposition permet de poser de nouvelles questions à propos des concepts humains et des processus cognitifs supérieurs. Finalement, je conclus que, une fois l'hypothèse EN abandonnée, la notion de concept perd l'essentiel de sa pertinence en psychologie cognitive.

# Chapitre 1

## Qu'est-ce qu'une théorie des concepts ?

---

Les concepts sont *des entités théoriques* introduites par les psychologues à des fins explicatives. Que sont ces concepts que les psychologues postulent ? Pourquoi introduire de telles entités théoriques ? Quel est l'objet d'une théorie des concepts ? Ce premier chapitre cherche à répondre à ces questions. L'objet de ce chapitre introductif est donc d'introduire la notion et de présenter, dans ses grandes lignes, la psychologie des concepts. Ce faisant, je souligne le contraste entre l'approche psychologique des concepts et l'approche philosophique. Je montre que philosophes et psychologues abordent la question des concepts de deux points de vue très différents. Je suggère d'embrasser l'approche psychologique : c'est à cette condition que de nouveaux problèmes philosophiques peuvent être formulés à propos des concepts.

Dans une première section, je présente la notion de concept utilisée par les psychologues des concepts. Dans la deuxième section, je présente trois hypothèses à propos de la classe des concepts : elles motivent et justifient l'introduction de cette notion théorique. Enfin, dans la dernière section, je décris l'objet des théories des concepts.

### 1. Qu'est-ce qu'un concept ?

#### 1.1. *Arrière-plan métaphysique*

Dans cette section, je présente les thèses métaphysiques qui sont implicitement admises par les théories psychologiques des concepts.

##### 1.1.1. Le réalisme intentionnel

Le réalisme intentionnel peut être défini comme suit.

## Le réalisme intentionnel

L'ontologie des psychologues cognitifs inclut des états mentaux<sup>1</sup> qui possèdent des propriétés sémantiques qui sont causalement explicatives.

Les réalistes intentionnels avancent trois thèses. Tout d'abord, les psychologues admettent l'existence d'états mentaux, c'est-à-dire d'entités qui sont différentes à la fois des stimuli physiques auxquels un organisme répond et de ses comportements, et qui, d'une manière ou une autre, assurent la médiation entre les uns et les autres. Ensuite, ces entités (ou, peut-être, certaines d'entre elles) peuvent être caractérisées sémantiquement : on peut leur attribuer des propriétés sémantiques, par exemple une référence, des conditions de vérité, des conditions de satisfaction etc. Remarquons que ces états mentaux ne sont pas nécessairement des croyances et des désirs. Tel qu'il est formulé ici, le réalisme intentionnel est neutre par rapport à la nature des états mentaux que possèdent les individus. Ainsi nos états mentaux ne sont pas nécessairement identiques aux états mentaux postulés par nos attributions mentalistes naïves. Enfin, ces états sont causalement explicatifs : la possession de ces états mentaux a une portée explicative.<sup>2</sup> Bien plus, ces états mentaux peuvent figurer dans des explications scientifiques des comportements des individus ainsi que de leurs processus cognitifs et la psychologie peut faire référence à ces états afin, par exemple, de formuler des lois et d'expliquer les comportements humains.<sup>3</sup>

Le réalisme intentionnel n'a pas toujours été accepté en psychologie. Particulièrement, les behavioristes ont cherché – sans succès – à élaborer une psychologie non mentaliste. Ils ont supposé qu'il était possible de formuler des théories psychologiques sans postuler l'existence d'états mentaux. De manière semblable, il est

---

<sup>1</sup> Dans tout ce qui suit, j'entends par « état mental » toute entité psychologique. Certains états mentaux font partie de notre psychologie naïve, par exemple les croyances, les désirs, les émotions, les humeurs etc. D'autres sont purement scientifiques comme les entités théoriques que sont la grammaire universelle, les prototypes etc.

<sup>2</sup> Ces trois thèses sont logiquement indépendantes. Stich (1983) accepte ainsi seulement la première.

<sup>3</sup> Je fais mienne la distinction proposée par Jackson et Pettit entre des propriétés *causalement efficaces* et des propriétés *causalement explicatives* (Jackson et Pettit 1988). Des propriétés causalement explicatives, par exemple des propriétés dispositionnelles, figurent dans des *explications causales* d'événements etc., bien que de telles explications supposent l'existence d'autres explications de ces mêmes événements dans lesquelles ces propriétés ne sont pas mentionnées. Puisque cette distinction n'est pas au cœur de ce travail, je l'admets sans la discuter.

possible que certains neurobiologistes contemporains cherchent aussi une description non mentaliste du comportement humain (par exemple, Staddon 2000). Nul ne sait aujourd'hui si cela peut être fait. En tout cas, les psychologues cognitifs contemporains acceptent le réalisme psychologique. Ils postulent l'existence d'états mentaux qui possèdent des propriétés sémantiques et qui expliquent causalement les comportements et processus cognitifs des agents.

*Les psychologues des concepts ne font pas exception.* Ce sont des réalistes intentionnels.

### **Concepts et réalisme intentionnel**

Les concepts sont supposés être des états mentaux qui ont des propriétés sémantiques et qui sont causalement explicatifs.

Les philosophes de la psychologie acceptent aussi cette position (par exemple, Fodor 1987, 1990, 1998, Laurence et Margolis 1999, Prinz 2002). Notons toutefois qu'il existe un débat philosophique sur la nature des concepts. Par exemple, de nombreux philosophes acceptent le réalisme intentionnel, mais nient que les concepts soient des états mentaux (par exemple, Peacocke 1992). Selon Peacocke, « concept » réfère à des entités *abstraites et normatives*. Les attitudes propositionnelles sont des états mentaux qui expriment des propositions, c'est-à-dire dont les valeurs sémantiques sont des propositions. Par exemple, la croyance de Jean que Marie est en retard exprime la proposition [Marie est en retard]. Les propositions sont des entités abstraites et structurées. Les concepts sont les constituants de ces entités abstraites. Par exemple, le concept ETRE EN RETARD est un composant de la proposition [Marie est en retard]. Pour exprimer une proposition, un penseur doit posséder les concepts qui constituent cette proposition (ces philosophes, souvent d'inspiration néo-frégéennes, disent habituellement « appréhender un concept », « to grasp a concept »). Et, il possède ces concepts seulement s'il possède certaines capacités. Par exemple, si quelqu'un ne connaît pas les règles d'introduction et d'élimination de la conjonction (ce savoir peut être pratique, et non théorique), il ne possède pas le concept ET, et ne pourrait donc croire que les chiens sont des animaux et les chaises des artefacts (Peacocke 1992).

Bien que cette position et les approches semblables (par exemple, Evans 1982) rejettent l'hypothèse de base de la psychologie des concepts, je ne vais pas discuter leur valeur. La principale raison est que les approches d'Evans, de Peacocke et des philosophes qui partagent leurs positions sont au fond des variations sur la question sémantique : comment un état mental a-t-il les propriétés sémantiques qu'il possède en fait ? Ces théories cherchent en fait à expliquer comment une attitude propositionnelle possède le contenu qui est le sien. Et je souhaite contourner ces questions. Je ne vais pas non plus défendre le réalisme intentionnel (voir, par exemple, Fodor 1987, 1990, Pacherie 1993, Jacob 1997 pour une discussion philosophique de cette position). Car si je m'engageais dans un tel travail, je ne n'en viendrais jamais à ce que je considère être les problèmes les plus intéressants. Au contraire, dans ce travail de thèse, j'adopte par stipulation l'usage des psychologues et de la plupart des philosophes de la psychologie. En conséquence, j'utilise « concept » pour référer à des états mentaux.

### 1.1.2. Représentationalisme

En sus du réalisme intentionnel, les psychologues admettent aussi le représentationalisme.

#### **Le représentationalisme**

Les états mentaux qui sont caractérisés sémantiquement et qui sont causalement explicatifs sont des représentations.

J'explique tout d'abord la notion de représentation. Une représentation est une entité particulière qui possède des propriétés sémantiques. Les propriétés intrinsèques non sémantiques de cette entité particulière caractérisent le *véhicule* de la représentation, tandis que les propriétés sémantiques sont le *contenu* de la représentation. Considérons par exemple l'énoncé suivant en anglais (en tant que token<sup>4</sup>) :

(1) Marie reads.

---

<sup>4</sup> Un token est une instance concrète. J'utilise {} pour nommer des types. Un type est une classe d'équivalence de tokens. « chien » et « chien » sont par exemple deux tokens du type {« chien »}. Deux états neuronaux dans deux individus peuvent être deux token concepts qui appartiennent au même type.

Cet énoncé possède à la fois des propriétés sémantiques et des propriétés intrinsèques non sémantiques, par exemple la forme et la taille des caractères. Par conséquent, (1) est une représentation. Pareillement, une peinture, par exemple la fameuse Corbeille de fruits du Caravage à Milan, est une entité particulière (qui peut être caractérisée en termes physiques) qui possède des propriétés sémantiques (ce qu'elle représente). C'est aussi une représentation. L'idée centrale du représentationalisme, c'est que les états mentaux que l'on peut caractériser sémantiquement sont des représentations précisément en ce sens.

La distinction entre le représentationalisme et le réalisme intentionnel est subtile. Cependant, on pourrait être un réaliste intentionnel sans accepter le représentationalisme. Par exemple, les deux thèses suivantes ne sont pas contradictoires :

- les états mentaux d'une personne donnée, à un temps  $t$ , surviennent<sup>5</sup> sur la constitution physique de son cerveau ;
- il est *impossible* d'isoler dans son cerveau une entité particulière qui serait pourvue des propriétés sémantiques de l'un de ses états mentaux. Les états mentaux surviennent *globalement* sur ces états neuronaux, bien qu'aucun état mental individuel ne survienne sur un état neuronal particulier.

La première de ces deux thèses correspond à l'acceptation d'un réalisme intentionnel matérialiste, tandis que la seconde nie le représentationalisme. Par conséquent, le représentationalisme et le réalisme intentionnel sont deux positions distinctes.

*Les psychologues des concepts endossent la thèse du représentationalisme.*

### **Concepts et représentationalisme**

Les concepts sont des représentations.

---

<sup>5</sup> En bref, la survenance (« supervenience ») entre les états mentaux et les états neuronaux signifie que deux individus qui ont les mêmes états neuronaux ont les mêmes états mentaux (voir, par exemple, Kim 1993, 1998). Plus généralement, un ensemble de propriétés A survient sur un ensemble de propriétés B si et seulement si il ne peut y avoir un changement dans A sans un changement dans B.

Tout concept en tant que token est un véhicule doué de certaines propriétés sémantiques. Par exemple, quand Jean a découvert l'existence de supernovas, une représentation SUPERNOVA s'est formée dans son cerveau.<sup>6</sup>

### 1.1.3. Conceptualisme

*Les psychologues des concepts admettent aussi le conceptualisme.*

#### **Le conceptualisme**

Certaines représentations font référence à des classes d'entités.<sup>7</sup>

La pensée que Marie lit a des conditions de vérité : elle est vraie si et seulement si Marie lit. Le conceptualisme est l'idée que toutes les représentations mentales ne sont pas de ce type. Certaines représentations *réfèrent* à des classes d'entités (par exemple, à des classes d'objets ou d'événements). Le conceptualisme affirme l'existence de représentations de ce genre. Les psychologues des concepts soutiennent que les concepts sont au nombre de ces représentations.<sup>8</sup> Notons que le conceptualisme n'exclut pas que ces représentations, par exemple les concepts, aient d'autres propriétés en sus de leur référence. Ainsi les concepts pourraient-ils posséder un sens en sus de leur référence.

## 1.2. *Qu'est-ce qu'un concept ?*

### 1.2.1. La notion de concept

Les psychologues admettent que les concepts sont des représentations mentales qui font référence à des classes d'entités. Mais toutes les représentations de ce genre ne sont pas des concepts, et une théorie psychologique des concepts porte sur un sous-ensemble de

---

<sup>6</sup> Notons que le fait que tout état mental, en tant que token, possède un véhicule n'implique en rien que les types d'états mentaux aient un véhicule spécifique. Car chaque état mental, en tant que token, pourrait avoir un véhicule distinct. Considérons un type de mots, par exemple {« chat »}. Il se peut, même si cela est discutable, qu'il n'y ait pas de propriété physique commune à tous les tokens de ce type. En conséquence, le fait qu'un token soit une représentation n'implique pas que le type correspondant ait aussi un véhicule spécifique.

<sup>7</sup> Dans ce travail, je ne distingue pas extension et référence. Habituellement, les termes singuliers, par exemple « Pierre », sont dits avoir une référence, tandis que les termes généraux, par exemple « chat », sont dits avoir une extension.

<sup>8</sup> On peut rejeter cette distinction en affirmant que les croyances font référence au vrai ou au faux.

ces représentations. L'objet de cette section est de caractériser plus précisément la notion de concept qui sera utilisée dans les chapitres suivants. Je ne *définis* pas la notion de concept, c'est-à-dire je ne propose pas un ensemble de conditions nécessaires et suffisantes pour qu'une représentation mentale (RM) soit un concept. L'idée est davantage d'indiquer quel type d'entités est l'objet d'une théorie des concepts en psychologie cognitive.

### **Les concepts et les fonctions supérieures**

Les concepts sont les représentations mentales à propos de classes d'entités qui sont utilisées par les capacités cognitives supérieures des êtres humains.

Expliquons d'abord certaines notions. Un organisme  $O$  possède une *capacité cognitive* ou une *compétence cognitive*  $\Gamma$  s'il satisfait les deux conditions suivantes :

1. Il y a deux classes d'états  $C_1$  (avoir tels stimuli, avoir telles représentations mentales etc.) et  $C_2$  (avoir telles représentations, faire telles actions etc.) telles que quand  $O$  est dans un état  $C_1$ , il passe dans un état  $C_2$ .  $C_1$  sont les inputs de  $\Gamma$  et  $C_2$  sont les outputs de  $\Gamma$ . Cette transition est appelée une *performance cognitive*.
2. Cette transition implique un ou plusieurs processus cognitifs.

*Un processus cognitif* est une manière spécifique de produire les outputs d'une capacité cognitive sur la base de ses inputs telle que :

- (i) ce processus contient des étapes intermédiaires ;
- (ii) chaque étape consiste à produire une représentation sur la base des représentations qui figurent dans les étapes antérieures ;
- (iii) le processus entier peut être décrit de manière sémantique, c'est-à-dire comme des inférences de propositions à partir d'autres propositions.

Je dis habituellement qu'une capacité cognitive est réalisée par un (ou plusieurs) processus cognitifs. Les représentations mentales figurent dans ces processus cognitifs et sont utilisées par les capacités cognitives correspondantes.

L'idée centrale de l'énoncé « Les concepts et les fonctions supérieures » est que les représentations mentales des humains peuvent être réparties dans plusieurs classes selon la nature des capacités qui les utilisent. Par exemple, les représentations perceptuelles des humains sont utilisées par défaut par nos capacités perceptuelles. Les représentations motrices sont utilisées par défaut par nos capacités motrices (voir, par exemple, Jacob et Jeannerod 2003). De même, les concepts sont les représentations qui sont utilisées par défaut par les capacités cognitives supérieures humaines.

Cette caractérisation réfère de manière un peu vague à *nos capacités cognitives supérieures*. Que sont-elles ? La distinction entre les capacités cognitives supérieures et les capacités cognitives inférieures ne peut être définie précisément. Ces notions ont peu de contenu théorique. Par capacité cognitive supérieure, j'entends simplement les capacités cognitives qui ont été traditionnellement, mais peut-être par erreur, associées avec la cognition des êtres humains. Ces capacités incluent avant tout la compréhension du langage, la catégorisation, le raisonnement déductif, le raisonnement inductif, la capacité analogique etc. Les capacités qui sont suffisamment semblables à ces capacités sont tenues pour des capacités cognitives supérieures.

La caractérisation de la notion de concept mentionne les capacités cognitives *des êtres humains*. Les représentations mentales qui sont utilisées par ces capacités sont les exemples caractéristiques des concepts. Cependant, cela ne revient en aucun cas à dénier par principe que d'autres espèces animales possèdent aussi des concepts. Si ces animaux ont des représentations qui sont suffisamment semblables aux représentations qui figurent dans les processus cognitifs supérieurs humains, nous n'avons aucune raison de nier qu'ils possèdent des concepts. Bien sûr, c'est une question empirique.<sup>9</sup>

---

<sup>9</sup> De ce point de vue, la sempiternelle question « Les animaux ont-ils des concepts ? » (voir par exemple Allen 1999) ne peut recevoir de réponse tant qu'un consensus à propos des représentations des êtres humains n'a pas émergé. Il y a peut-être cependant une autre approche. Supposons qu'une espèce animale quelconque possède des capacités cognitives qui soient semblables à certaines de nos fonctions cognitives supérieures. Alors, nous pourrions étudier de manière empirique la nature des représentations qui sont utilisées par ces capacités. Nous pourrions les appeler « concepts animaux », ou, peut-être « concepts

### 1.2.2. Encodage des connaissances

Les psychologues supposent que les concepts *encodent certaines connaissances* à propos de ce qu'ils dénotent. Un concept est *un ensemble de connaissances à propos d'une classe particulière*.<sup>10</sup> Par exemple, un concept de chien est supposé encoder certaines connaissances à propos des chiens.<sup>11</sup> On peut caractériser les connaissances qui sont encodées par un concept comme un ensemble de propositions à propos de ce que ce concept dénote. Par exemple, un concept de chien exprime un ensemble de propositions que l'agent cognitif tient pour vraies de la classe des chiens. Remarquons que les propositions ne sont pas par essence exprimables linguistiquement : nos langues naturelles peuvent ne pas avoir les ressources sémantiques pour décrire les propositions exprimées par les concepts. Certaines propositions peuvent être linguistiquement exprimables, par exemple [les chiens aboient], d'autres non. C'est une question empirique de savoir si les ensembles de connaissances qui sont encodés par les concepts peuvent être linguistiquement exprimés.<sup>12</sup>

Dans quelles conditions un état neuronal exprime-t-il un ensemble de propositions ? Comme mon propos n'est pas sémantique, je ne présente pas en détail une théorie psychosémantique. Je soutiens toutefois qu'un concept exprime un ensemble de propositions si la condition suivante est satisfaite : *si nous lui attribuons ce contenu, nous pouvons décrire les processus dans lesquels il figure en termes sémantiques* (voir Cummins 1989, chap. 8, pour une idée semblable). On veut dire par là que ce concept figure dans des processus cognitifs qui ont les caractéristiques suivantes :

---

canins », « concepts félins » etc., qu'elles ressemblent ou non aux concepts humains. Et nous pourrions alors les comparer à nos concepts.

<sup>10</sup> Dans ce travail, comme très souvent en psychologie, la notion de *connaissance* n'est pas ici prise en son sens philosophique. Elle est plus ou moins identique à la notion de croyance. En outre, la notion d'*encodage* n'est pas développée en détail dans ce travail (mais voir la sémantique minimaliste présentée brièvement ci-dessous).

<sup>11</sup> On pourrait utiliser la terminologie philosophique habituelle. L'ensemble de connaissances qui est encodé par un concept pourrait être appelé son *sens* ou son *contenu cognitif*. Cependant, puisque ces notions sont liées à des problèmes spécifiquement philosophiques, je préfère ne pas les utiliser.

<sup>12</sup> De nombreuses données empiriques suggèrent qu'il est difficile d'exprimer linguistiquement les connaissances qui sont encodées par les concepts. Cela est suggéré par la recherche sur l'apprentissage implicite (voir, par exemple, Shanks 1997b et Knowlton 1997) et par les modèles connexionnistes des concepts (voir, par exemple, Clark 1993).

- (i) Ils peuvent être décrits de manière sémantique : les états successifs de ces processus cognitifs peuvent être caractérisés comme exprimant certaines propositions et les transitions entre ces états comme des inférences.
- (ii) L'occurrence du concept C au sein de ces processus cognitifs peut être décrite en termes des propositions que C est supposé encoder.

Bref, je suggère d'accepter la sémantique suivante :

### **Sémantique minimaliste**

Un concept exprime la proposition p si, lorsque nous lui attribuons p, nous pouvons décrire en termes sémantiques les processus dans lesquels il figure.<sup>13</sup>

#### 1.2.3. Le critère de démarcation et le problème du holisme

Les connaissances qui sont encodées par un concept constituent un sous-ensemble de toutes nos connaissances à propos de ce que ce concept représente. Considérons par exemple un concept GRAND-MERE. Supposons que ce concept exprime la définition [x est une grand-mère si et seulement si x est la mère d'un enfant qui a aussi un enfant]. Supposons aussi que Marie a des croyances idiosyncrasiques à propos des grand-mères : par exemple, elle croit que les grand-mères deviennent des loups la nuit. Cependant, par hypothèse, cette croyance, tout comme de nombreuses autres, n'est pas représentée par son concept de grand-mère. Cela soulève naturellement la question suivante : *comment distingue-t-on les propositions qui sont exprimées par un concept de celles qui ne le sont pas ? Il s'agit du problème du critère de démarcation.*<sup>14</sup>

Certains philosophes et psychologues adoptent *une réponse holiste* (Quine 1953, Carey 1985, 1991, Carey et Johnson 2000, Pacherie 1997, Churchland 1998...) : le concept que possède un individu x encode toutes les connaissances de x à propos de ce à quoi le concept réfère. Ainsi le concept holiste GRAND-MERE possédé par Marie

<sup>13</sup> Clairement, cette sémantique devrait être élaborée en détails pour pouvoir rivaliser avec les sémantiques actuelles. Comme je tiens à éviter les questions sémantiques au profit des questions de philosophie des sciences, je me contente de la formulation présentée ci-dessus. Remarquons cependant qu'elle est proche des idées sémantiques de nombreux philosophes (Dennett 1987, Cummins 1989, Field 2001).

<sup>14</sup> Fodor a justement reproché aux psychologues de ne pas fournir un tel critère (Fodor 1998).

exprime-t-il, entre autres, la proposition selon laquelle les grand-mères se transforment en loups la nuit.

Mais cette théorie des concepts n'est pas satisfaisante. De nombreux philosophes ont fait l'objection suivante : si cette forme de holisme était vraie, deux individus ne pourraient avoir le même concept (Fodor et Lepore 1992, Lepore et Fodor 1999). Car les individus possèdent des croyances différentes, entre autres parce que leurs expériences sont différentes. En réponse, on concède habituellement que deux personnes ne peuvent avoir des concepts *identiques*. Mais, poursuit-on, elles peuvent avoir des concepts *semblables* (par exemple, Churchland 1998). Fodor a critiqué à de nombreuses reprises cette réponse (Fodor et Lepore 1992, Fodor 1998, Lepore et Fodor 1999). Il soutient à juste titre que deux concepts sont semblables si et seulement s'ils encodent les mêmes propositions. Or, affirme-t-il, deux concepts expriment la même proposition  $\Gamma$  seulement s'ils sont constitués à partir des mêmes concepts : malgré leur différences, les concepts de chien possédés par deux individus A et B expriment la même proposition [les chiens aboient], seulement si ces deux concepts sont composés à partir du *même* concept ABOYER. En conséquence, la *ressemblance* entre concepts suppose que différents individus puissent avoir des concepts *identiques* – ce que le holisme semble rendre impossible.

La réponse de Fodor suppose l'idée suivante : un état mental exprime une proposition si et seulement si c'est une pensée complexe faite à partir des concepts qui correspondent aux composants de cette proposition.<sup>15</sup> Pour exprimer la proposition [les chiens aboient], on doit avoir un état mental composé à partir de CHIEN et ABOYER. Comme nous l'avons vu précédemment, je soutiens au contraire qu'un état mental exprime une proposition si l'attribution de cette proposition à cet état permet de décrire de manière sémantique les processus dans lesquels il figure. En conséquence, la ressemblance des concepts ne suppose pas leur identité. Quand nous disons que les concepts de chien possédés par deux individus A et B sont semblables, nous voulons bien dire qu'ils expriment des ensembles semblables de propositions. Et cela est le cas, parce que le concept de chien possédé par A et le concept de chien possédé par B figurent dans des processus semblables : A et B catégorisent les mêmes chiens comme étant des chiens,

---

<sup>15</sup> Chez Fodor, cette idée prend la forme spécifique du langage de la pensée (Fodor 1975, 1981).

ils font *grosso modo* les mêmes inférences inductives à propos des chiens etc. Bref, pour que ces deux concepts soient semblables, il n'est nul besoin de supposer qu'ils soient constitués de concepts identiques. Je n'endorsse donc pas la critique du holisme par Fodor.

Pourtant, le holisme ne peut être une théorie des concepts. Le vrai problème du holisme est qu'il est incompatible avec l'idée même de concept. En effet, un holiste soutient que lorsqu'un individu est impliqué dans une tâche de catégorisation, de déduction ou d'induction, par exemple à propos des chiens, il ne s'appuie pas de manière préférentielle sur un ensemble spécifique de connaissances à propos des chiens.<sup>16</sup> Nier l'existence d'un ensemble spécifique de connaissances qui soit *par défaut* impliqué dans ces performances cognitives revient à se débarrasser en fait de la notion psychologique de concept.

Comment donc distinguer les connaissances qui sont représentées par le concept d'une classe de celles qui ne le sont pas ? Je propose la réponse suivante.

### **Le critère de démarcation**

Les connaissances qui sont constitutives d'un concept sont *activées* (c'est-à-dire, extraites de la MLT) *par défaut* dans nos processus cognitifs supérieurs à propos de ce qu'il représente.

Ainsi la croyance de Marie que les grand-mères se transforment en loups la nuit ne fait-elle pas partie de son concept de grand-mère puisqu'elle n'est pas activée par défaut quand Marie est impliquée dans une tâche cognitive à propos des grand-mères. Au contraire, la connaissance que x est une grand-mère si et seulement si x est la mère d'un enfant qui a aussi un enfant est activée par défaut. Considérons aussi le cas du concept de

---

<sup>16</sup> Bizarrement, Fodor est d'accord avec les holistes sur ce point (Fodor 1983). Il rejette par exemple l'idée que nos performances de catégorisation, par exemple dans la classe des chiens, reposent par défaut sur un ensemble spécifique de connaissances à propos des chiens. Selon lui, nous utilisons n'importe quel type de connaissances afin de prendre une décision de catégorisation (voir aussi Rey 1983, 1985). Il est de toute évidence vrai que nous *pouvons* utiliser n'importe quel type de connaissances à propos des chiens quand nous catégorisons au sein de la classe des chiens, quand nous raisonnons inductivement à leur propos, quand nous faisons des analogies avec les chiens... Cela n'implique cependant en rien que nous ne fassions pas appel *par défaut* à un ensemble spécifique de connaissances.

chien de Paul. Quand Paul pense aux chiens (raisonne inductivement ou déductivement à leur propos, catégorise un objet comme étant un chien, fait une analogie avec les chiens...), les connaissances qu'il extrait par défaut de la MLT constituent son concept CHIEN.

#### 1.2.4. L'objection méréologique

On trouve dans la littérature philosophique l'objection suivante contre l'idée que les concepts expriment des ensembles de propositions. Exprimer une proposition à propos des  $x$ , c'est avoir une pensée à propos des  $x$  et les pensées à propos des  $x$  sont composées à partir du concept  $X$ . Un concept de  $x$  ne peut donc exprimer un ensemble de propositions à propos des  $x$ .<sup>17</sup> Par exemple, le concept de chien ne peut exprimer la proposition que les chiens ont des poils. Car exprimer cette proposition, c'est avoir une pensée que les chiens ont des poils. Mais cette pensée est un état complexe fait entre autres à partir du concept CHIEN. En conséquence, CHIEN n'exprime pas cette proposition, mais est un constituant de la pensée qui exprime cette proposition. Généralement, comment un concept pourrait-il exprimer un ensemble de propositions si exprimer ces propositions consiste à avoir des états mentaux qui sont composés entre autres à partir de ce concept ?

L'étape centrale de ce raisonnement est l'affirmation qu'exprimer une proposition, par exemple, [les chiens sont poilus], c'est avoir un état mental composé par les concepts qui correspondent aux composants de cette proposition, dans le cas présent par les concepts CHIEN et POILU. Cependant, comme nous l'avons vu ci-dessus, je rejette cette approche du contenu des concepts. Soutenir qu'un concept est un ensemble de connaissances, c'est affirmer que si nous lui attribuons ce contenu, nous pouvons interpréter sémantiquement les processus dans lesquels il figure. Un concept peut donc exprimer une proposition sans être lui-même composé. Par conséquent, l'objection méréologique n'est pas décisive. Elle suppose que les concepts expriment une proposition seulement s'ils sont composés. Mais cette hypothèse n'est nullement nécessaire une fois que l'on admet la sémantique minimaliste proposée ci-dessus.

---

<sup>17</sup> Voir, par exemple, Laurence et Margolis 1999 : les concepts ne peuvent être des théories, car les théories sont faites de concepts. Cette objection peut être généralisée contre toute conception selon laquelle les concepts encodent des ensembles de connaissances.

### *1.3. Comparaison avec la caractérisation philosophique de la notion de concept*

La manière dont j'ai jusqu'à présent caractérisé la notion de concept diffère grandement de l'approche philosophique habituelle. Comme nous l'avons vu à plusieurs reprises, les philosophes partent en général de l'idée suivante : les concepts sont les constituants de nos pensées (par exemple, Evans 1982, Rey 1983, 1985, Fodor 1998, Laurence et Margolis 1999, Prinz 2002).<sup>18</sup> Les êtres humains, et peut-être d'autres animaux, ont des attitudes propositionnelles, par exemple des croyances et des désirs. Ces états mentaux sont des représentations mentales complexes (voir Cummins 1989, pour une discussion critique de cette idée).<sup>19</sup> Par exemple, la croyance de Pierre que Marie est en retard est une représentation composée d'autres représentations. Les concepts sont les représentations mentales dont, en un sens habituellement non spécifié, ces attitudes propositionnelles sont faites. Par exemple, le concept ETRE EN RETARD est un constituant de la croyance de Pierre.

L'idée que les concepts sont les constituants de pensées complexes ne joue aucun rôle dans mon approche. Au contraire, j'ai exprimé à plusieurs reprises mes doutes sur cette idée. La sémantique minimaliste que j'ai esquissée implique qu'un état mental peut exprimer des propositions complexes sans être lui-même composé. Inversement, l'idée que les concepts figurent dans des processus cognitifs supérieurs n'est pas au cœur de l'approche philosophique. C'est au contraire mon point de départ.

L'approche que je propose s'accorde avec la manière dont les psychologues conçoivent les concepts. Selon eux, les concepts sont des ensembles de connaissances à propos de classes d'entités qui sont utilisés par défaut par certaines capacités psychologiques, en premier lieu la catégorisation, le raisonnement inductif et le raisonnement déductif. Par exemple, un concept de chien est un ensemble de

---

<sup>18</sup> La nature de cette relation de constitution est habituellement non explicitée (Machery 2002). L'hypothèse du langage de la pensée est une manière de l'expliciter (par exemple, Fodor 1975, Carruthers 1996, Devitt et Sterelny 1999). Selon cette hypothèse, les concepts sont des ensembles de nos pensées des énoncés formés à partir de ces mots. Cependant, les philosophes et les chercheurs en sciences cognitives qui acceptent l'idée de constitution n'admettent pas toujours le langage de la pensée (par exemple, Evans 1982, Prinz 2002).

<sup>19</sup> Certains philosophes tiennent cette thèse pour empirique : c'est la meilleure explication de la productivité et de la systématisme de notre pensée (Fodor 1998, Fodor et Pylyshyn 1988, Fodor et McLaughlin 1990). D'autres soutiennent que cette thèse est présupposée par l'attribution naïve d'états mentaux : nous n'attribuerions pas de croyances à une créature dont les états mentaux ne seraient pas complexes (Evans 1982, Davies 1991).

connaissances à propos des chiens qui figure par défaut dans nos catégorisations, nos déductions à propos des chiens, nos inductions à propos des chiens etc. Les psychologues admettent l'existence de ces ensembles de connaissances et la psychologie des concepts entreprend de les caractériser.<sup>20</sup> Ainsi Barsalou écrit-il (2003b, 84) :

« The human conceptual system contains people's knowledge of the world. In most theories, the basic unit of knowledge is the concept. (...) this construct is highly contentious. Following psychological theories, we assume that a concept, roughly speaking, is knowledge about a particular category (e.g., *birds*, *eating*, *happiness*). Thus, knowledge about *birds* represent the bodies, behaviors, and origins of the respective entities. Knowledge plays central roles throughout the spectrum of cognitive activities. In on line processing of the environment, knowledge guides perception, categorization, and inference. In off line processing of non present situations, knowledge reconstructs memories, underlies the meanings of linguistic expressions, and provides the representations manipulated in thought. »

Certains psychologues des concepts admettent aussi que les concepts sont les constituants des pensées (voir, par exemple, Osherson et Smith 1981, Armstrong et al. 1983, Murphy 2002, 443-444). *Cependant, cette thèse ne joue habituellement qu'un rôle très limité dans la psychologie des concepts* (Barsalou 1999 est une exception). Ainsi, après avoir soutenu, au début du chapitre sur la composition des concepts, que les concepts sont les composants de nos pensées, Murphy affirme qu'il ne discutera pas la construction des pensées, mais seulement des concepts complexes comme CHIEN FOU ou ETUDIANT D'OXFORD QUI EST UN MENUISIER (Murphy 2002, chap. 7). De

---

<sup>20</sup> Barsalou et al. écrivent (2003b, 84) : « Cognitive scientists continue to argue about the construct of *concept*. For example, Platonists propose that concepts exist independently of minds, whereas naturalists argue that concepts are causal relations between the world and the mind (...). From these two perspectives, the study of concepts constitutes a referential problem, namely, how do mental states refer to categories in the world ? Many other cognitive scientists – especially psychologists – focus on the cognitive side of this relation, assessing the structure, format, and content of conceptual representations. » J'approuve cette distinction. Et je crois que l'aspect cognitif de cette relation soulève des questions philosophiquement importantes.

fait, l'idée que les concepts sont les constituants des pensées est entièrement absente du chapitre.

*Je propose de suivre l'exemple des psychologues et de négliger l'idée de pensées faites à partir de concepts.* La principale raison pour ne pas adopter l'approche philosophique est pragmatique : se représenter les concepts comme des ensembles de connaissances à propos de classes d'entités qui sont par défaut utilisées par nos compétences cognitives supérieures, comme le font les psychologues, permet de poser des questions importantes et philosophiquement intéressantes. Je n'ai pas l'intention d'affirmer, arguments à l'appui, que l'approche philosophique est erronée. Je la mets simplement en parenthèse. Cette stratégie un peu cavalière sera justifiée si la discussion des théories psychologiques se révèle philosophiquement productive.

#### *1.4. Conclusion*

Dans cette section, j'ai introduit la notion psychologique de concept. Les psychologues admettent l'existence d'états mentaux qui sont caractérisés sémantiquement et qui ont une portée explicative. Ces états mentaux sont des représentations. Certaines de ces représentations font référence à des classes d'entités. Parmi celles-ci, on trouve les concepts : ce sont les représentations qui figurent dans les processus cognitifs supérieurs. Un concept à propos des  $x$  encode les connaissances qui sont extraites par défaut de la MLT quand son possesseur s'engage dans une tâche à propos des  $x$  qui dépend d'une faculté cognitive supérieure. Les objections philosophiques contre cette approche reposent sur l'idée que les concepts ne peuvent exprimer une proposition que s'ils sont complexes : par exemple, le concept GRAND-MERE ne peut exprimer la proposition [les grand-mères sont les mères de parents] que s'il est composé des concepts MERE et PARENT. J'ai rejeté cette idée en admettant une sémantique minimaliste. Dans ce travail, j'utilise donc la notion psychologique de concept.

## **2. Hypothèses théoriques**

### *2.1. Hypothèses et signification*

La plupart des notions théoriques scientifiques sont associées à des *hypothèses théoriques*. Ces hypothèses *motivent* l'introduction de ces notions. Par exemple, la croyance de Newton que les corps matériels ont une propriété qui est une fonction de leur quantité de matière et que cette propriété est liée nomologiquement à certaines autres propriétés, par exemple à la vitesse de certains mouvements, motive l'introduction de la notion théorique de masse.

Les philosophes ont souvent présenté cette idée en termes *sémantiques*. La question « est-ce que des propositions théoriques constituent *la signification* d'un terme théorique ? » a en effet suscité de nombreux débats philosophiques (Kuhn 1970, Lewis 1979, Feyerabend 1975, 1981, Putnam 1975, Kitcher 1993). Par exemple, on peut se demander si les propositions [les corps ont une grandeur qui est proportionnelle à leur quantité de matière] et [cette quantité est nomologiquement liée à leur vitesse en chute libre] sont constitutives de la signification du terme théorique « masse » (Field 1973). Certains philosophes ont défendu une position holiste : *tous* les énoncés scientifiques dans lesquels un terme scientifique figure définissent sa signification (Kuhn 1970, Lewis 1979). Au contraire, d'autres ont défendu un atomisme radical, en niant que la signification d'un terme théorique soit constituée par la moindre proposition théorique (Putnam 1975). La plupart des philosophes semblent préférer une position intermédiaire, selon laquelle seules certaines propositions constituent la signification d'un terme scientifique. Le problème est délicat, dans la mesure où il fait appel à la distinction entre le synthétique et l'analytique et à l'incommensurabilité de la signification.

On le voit, une telle approche est prise dans les controverses sémantiques. Cela est problématique, étant donné qu'aucun consensus n'a émergé en sémantique et que la possibilité d'un tel consensus est même douteuse (Stich 1996, Bishop et Stich 1999, Machery et al. 2004). Il n'est pas question de développer ici un cadre non sémantique pour traiter ces questions. Contentons-nous plutôt de la formulation un peu vague utilisée au début de cette section : certaines propositions *motivent* l'introduction de notions scientifiques. Dans ce qui suit, je présente les hypothèses principales qui sont associées avec la notion de concept.

## 2.2. Promiscuité et promiscuité renforcée

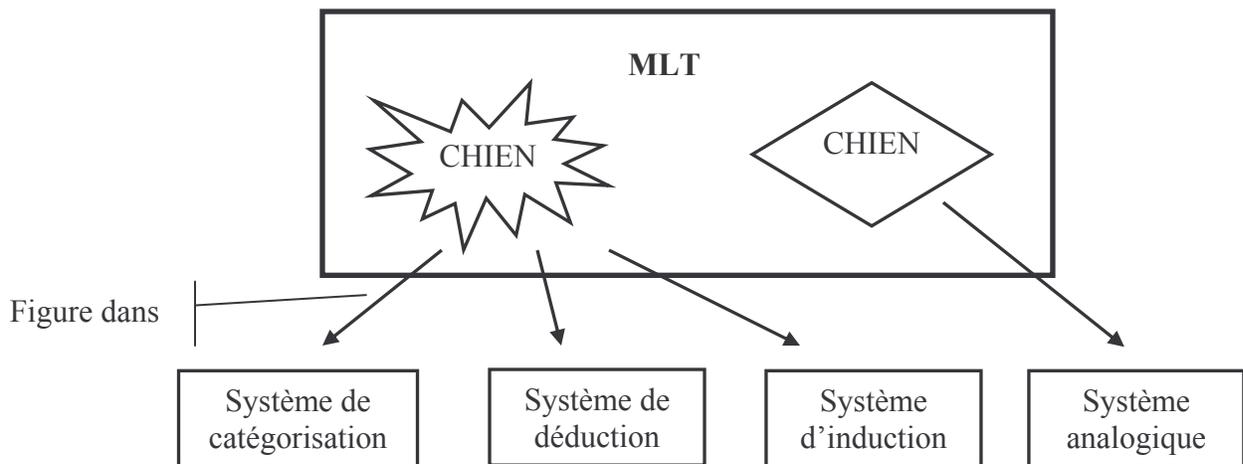
### 2.2.1. Le principe de promiscuité

La première hypothèse concerne l'occurrence des concepts dans les processus cognitifs.

#### **Le principe de promiscuité**

Un concept est utilisé par de nombreuses compétences cognitives supérieures.<sup>21</sup>

Par exemple, quand Jean catégorise un objet comme étant un chien, raisonne inductivement ou déductivement à propos des chiens etc., les processus cognitifs dont dépendent ces performances impliquent le même concept de chien. La position contraire soutiendrait que *différents types* de représentations figurent dans les processus cognitifs qui réalisent différentes compétences cognitives supérieures. Par exemple, la représentation qui figure dans les catégorisations de Jean serait différente de la représentation qui figure dans ses raisonnements déductifs etc. (voir le chapitre 2, section 2.3.a.). Si cela était le cas, la notion de concept perdrait beaucoup de sa pertinence. Pour exprimer l'idée que les concepts figurent dans de nombreux processus, je dis que les concepts ont la propriété de promiscuité. Le schéma suivant exprime cette hypothèse.



**Schéma 1.1.**

<sup>21</sup> Voir la notion de promiscuité inférentielle dans Stich 1978 (« inferential promiscuity »). Selon Stich, les croyances ont la propriété de promiscuité inférentielle parce qu'elles figurent dans de nombreux raisonnements.

Dans cet exemple fictif, plusieurs compétences cognitives, mais pas toutes, font appel à un même concept de chien. La capacité de faire des analogies fait appel à une autre représentation de chien.

Cette hypothèse est admise par *tous* les psychologues et les philosophes qui croient que la notion de concept est théoriquement utile. Les citations suivantes illustrent son acceptation par d'importants psychologues<sup>22</sup> :

*« the same packet of information is employed in a wide range of contexts, including the case where a concept is used in combination with other concepts. »*  
Medin et Shoben 1988, 158.

*« Concepts function in enormously varied ways. They can be used for extremely rapid identification (as when escaping from prey), organizing information efficiently in memory, problem-solving, analogizing, drawing inductive inferences, that extend knowledge beyond what is known, embodying and imparting ideological inferences, conveying aesthetic materials (e.g., metaphor, poetry), and so forth. These are different kinds of judgments that partake of different kinds of information. People rely strongly on perceptual similarity with some sorts of judgments, but focus on perceptual features more with others. One might expect beliefs about essences to figure more prominently in judgments associated with Frank Keil's (1989) transformation experiments than in run-of-the-mill object recognition. In short, conceptual functions go beyond categorization.*

(...) different investigators focus on different conceptual functions, and often researchers fail to clarify which function is under study. Jones and Smith stress the identification function of concepts. In fact, they define « concept » as the represented structure « that allows members of a category to be recognized. » We find this definition overly narrow. » Gelman et Medin 1993, 158-59.<sup>23</sup>

---

<sup>22</sup> Je souligne.

<sup>23</sup> Voir aussi Gelman et Coley 1991, 162, Murphy 2002, 3.

### 2.2.2. Validité empirique du principe de promiscuité

Cette hypothèse est validée par un large ensemble de données. Par exemple, Gopnik et Meltzoff ont présenté de nombreuses données concernant deux aspects du développement cognitif des enfants (Gopnik et Meltzoff 1997) :

- *Le développement conceptuel des enfants varie d'un domaine à l'autre (« domain-specific »)* : le développement des concepts des enfants varie, par exemple, entre la psychologie naïve et la théorie naïve des objets. Cette idée est largement acceptée par les psychologues développementaux contemporains (voir, par exemple, Carey 1985, Perner 1991, et les nombreux articles dans Hirschfeld et Gelman 1994 et Medin et Atran 1999).
- *Mais le développement conceptuel ne varie pas d'un processus à l'autre (« process-specific »)* : pour un domaine donné, par exemple la théorie naïve des objets, le développement des concepts des enfants se manifeste dans des tâches variées qui requièrent différents processus cognitifs. Cela suggère que le même concept figure dans ces différents processus.

Il y a aussi des données empiriques convaincantes en faveur du principe de promiscuité pour la cognition adulte. On trouve en effet les mêmes phénomènes dans des tâches expérimentales qui reposent sur différentes compétences cognitives. Par exemple, nos performances de catégorisation sont extrêmement sensibles à la typicalité des objets catégorisés (Smith et al. 1974, Rosch 1978). Nos raisonnements inductifs le sont aussi (Smith 1989, Osherson et al. 1990), tout comme nos raisonnements déductifs (Sloman 1998).<sup>24</sup>

### 2.2.3. Le principe de la promiscuité renforcée

Remarquons finalement que le principe de promiscuité est différent du principe suivant :

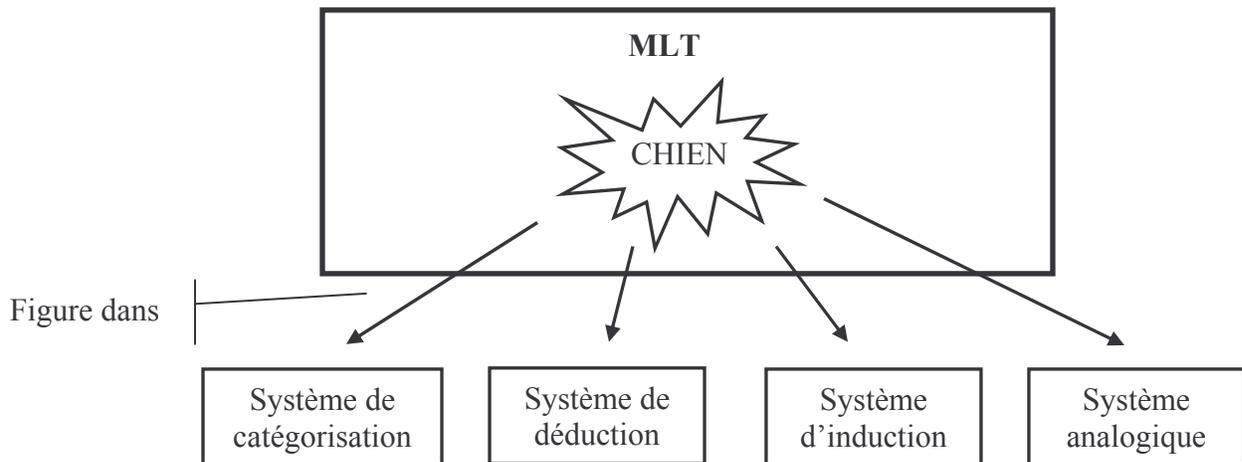
#### **Le principe de la promiscuité renforcée**

Un concept est utilisé par *toutes* les compétences cognitives supérieures.

---

<sup>24</sup> Voir le chapitre 4, section 2.2.

L'idée est que *toutes* nos performances cognitives supérieures à propos, par exemple, des chiens reposent sur un unique concept, CHIEN. Le principe de promiscuité affirme seulement qu'un concept est utilisé par *de nombreuses* compétences cognitives supérieures. Le principe de la promiscuité renforcée est décrit par le schéma suivant :



**Schéma 1.2.**

Les psychologues ne discutent habituellement pas cette version forte de la promiscuité. Il se peut que certains admettent cette version, mais il est difficile de le déterminer. Certains philosophes peuvent aussi accepter ce principe. Par exemple, Fodor semble adhérer au principe de la promiscuité renforcée : un processus conceptuel est à propos des chiens seulement s'il implique des croyances à propos des chiens, c'est-à-dire des croyances qui sont faites à partir du concept CHIEN (sauf si le sujet a des concepts coréférentiels de chiens<sup>25</sup>).

Les données en faveur du principe de promiscuité sont bien sûr cohérentes avec le principe de la promiscuité renforcée, sans le confirmer davantage que le principe de promiscuité. Comme le principe de la promiscuité renforcée est difficile à établir empiriquement et comme il n'est pas évident qu'il a des partisans, je propose que la notion de concept repose seulement sur le principe de promiscuité.

### 2.3. L'efficacité causale des concepts

<sup>25</sup> Deux concepts sont coréférentiels ou coréfèrent s'ils font référence à la même classe.

### 2.3.1. Le principe de l'efficacité causale des concepts

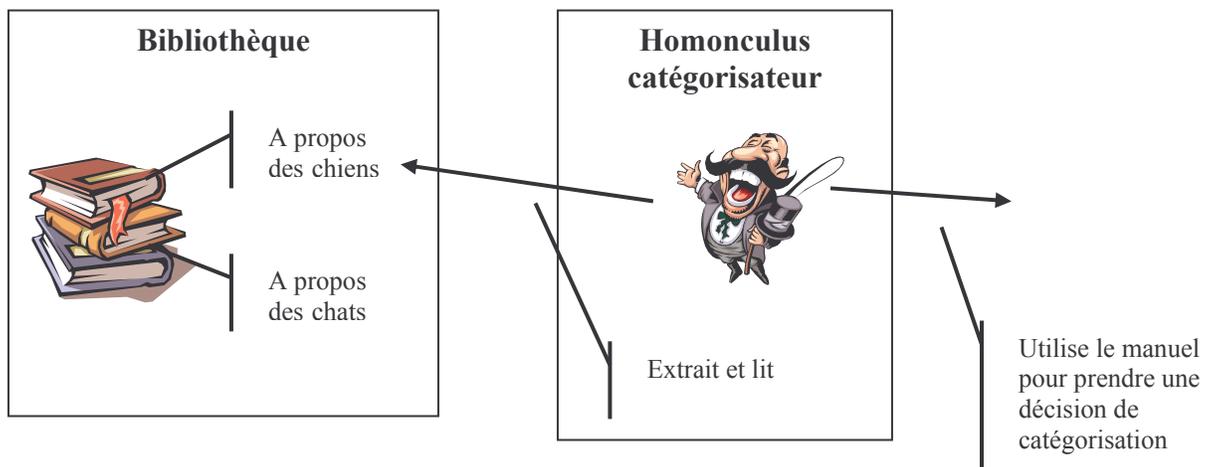
La deuxième hypothèse concerne la manière dont les concepts figurent dans les processus cognitifs.

#### **Le principe de l'efficacité causale**

Les propriétés des concepts contribuent à expliquer causalement le déroulement des processus cognitifs dans lesquels ils figurent.

Par exemple, la manière dont Marie catégorise Médor comme un chien est causalement expliquée par les propriétés de son concept CHIEN. Si elle avait un autre concept de chien, par exemple CHIEN<sub>2</sub>, elle ne catégoriserait pas Médor comme elle le fait.

Nous pouvons donc traiter chaque processus cognitif, par exemple le processus dont dépend la catégorisation<sup>26</sup>, comme un homonculus qui est en charge de satisfaire la fonction du processus cognitif. Pour catégoriser un objet dans une catégorie donnée, par exemple la catégorie des chiens, l'homonculus catégorisateur extrait d'une bibliothèque (c'est-à-dire, de la MLT) un manuel qui contient un ensemble de connaissances à propos des chiens (c'est-à-dire, le concept). Il utilise cette information pour décider si l'entité appartient à la catégorie des chiens. S'il doit catégoriser un objet dans la catégorie des chats, il extrait un autre manuel.



**Schéma 1.3.**

<sup>26</sup> Je suppose ici provisoirement que ce processus est unique. Je critique cette hypothèse dans le chapitre 5.

Cette idée est admise par *presque tous* les psychologues contemporains (à l'exception, peut-être, de Gleitman). Elle est au fondement de *l'association étroite entre l'étude des concepts* (quelles sont les propriétés des concepts ?) et *l'étude des processus conceptuels*. Ces recherches se contraignent l'une l'autre : puisque le déroulement des processus cognitifs supérieurs est supposé être causalement expliqué, entre autres, par les propriétés des concepts, ce que les psychologues savent à propos des concepts et ce qu'ils savent à propos des processus conceptuels doivent être intimement liés. Par exemple, de nombreuses théories psychologiques des concepts (particulièrement, de nombreuses théories des prototypes et de nombreuses théories des exemplaires<sup>27</sup>) sont associées à des modèles détaillés des processus dans lesquels ils sont supposés figurer, en particulier à des modèles détaillés de la catégorisation (par exemple, Medin et Schaffer 1978, Nosofsky 1986, 1988, Ross et Makin 1999, voir le chapitre 5, section 2.).

Le principe de l'efficacité causale est aussi admis par de nombreux philosophes de la psychologie. Clark (1993), Laurence et Margolis (1999), Gärdenfors (2000) et Prinz (2002) admettent que les théories des concepts doivent expliquer nos performances de catégorisation (voir, en particulier, Clark et Prinz 2004).

### 2.3.2. Commentaires

Les philosophes et les psychologues ne devraient pas insister exclusivement sur la catégorisation (voir la citation de Gelman et Medin ci-dessus) : les concepts figurent dans un grand nombre de processus cognitifs supérieurs (voir le principe de promiscuité). Bien que la catégorisation soit de toute évidence un processus cognitif supérieur crucial (voir le chapitre 5, section 1.), de nombreux psychologues, par exemple Susan Gelman, ont à juste titre souligné qu'il importe de prendre en considération la totalité des processus dans lesquels figurent nos concepts. Une théorie des concepts devrait être en mesure d'expliquer les données associées avec nos compétences cognitives supérieures – ou au moins avec la plupart d'entre elles (voir le principe de promiscuité). En conséquence, elle doit être testée avec des données provenant de l'étude des différents processus cognitifs.

---

<sup>27</sup> Voir le chapitre 3, section 1.

Le principe de l'efficacité causale explique pourquoi les théories des concepts sont jugées si importantes pour comprendre l'esprit. Avoir une telle théorie est nécessaire pour expliquer correctement nos capacités cognitives : on ne peut comprendre un processus cognitif si l'on néglige les représentations qu'il utilise. Cela explique sans doute pourquoi le débat à propos des concepts a été souvent si virulent (voir l'échange entre Fodor 2000 et Hampton 2000).

### 2.3.3. Défense du principe de l'efficacité causale

Le principe de l'efficacité causale n'est pas universellement admis. L'un des principaux théoriciens des concepts, Jerry Fodor, rejette l'idée que les propriétés des concepts expliquent les propriétés des processus conceptuels dans lesquels ils figurent (Fodor 1998, 2004a, b). Fodor admet l'approche philosophique des concepts : les concepts sont les constituants des pensées. Sur cette base, il soutient que les *croyances* des individus expliquent le cours des processus cognitifs : Marie catégorise Médor comme un chien parce qu'elle *croit* que les chiens ont telle et telle propriété et qu'elle voit que Médor a ces propriétés. Comment répondre à cette objection ?

Notons tout d'abord les deux faits suivants. D'un point de vue sociologique, la position de Fodor est *extrêmement marginale* dans le champ des concepts – bien qu'il ait davantage de défenseurs en philosophie qu'en psychologie. En outre, elle est *extrêmement révisionniste*. Fodor ignore simplement l'une des motivations traditionnelles pour postuler des concepts : philosophes et psychologues ont postulé l'existence de concepts en partie pour expliquer certains de nos processus conceptuels – traditionnellement, la catégorisation et l'inférence. Ces deux points n'impliquent toutefois évidemment pas que la position de Fodor est fausse.

On peut avoir l'impression que les concepts de Fodor ne jouent aucun rôle explicatif. Cela serait une objection décisive. Pourquoi en effet poser ces entités théoriques si elles ne jouent aucun rôle explicatif ? Cependant, les concepts ont bien un rôle explicatif dans l'approche de Fodor. Ils sont supposés être les éléments d'un système combinatoire de représentations. Selon Fodor, l'existence d'un tel système est la meilleure explication de deux propriétés essentielles de notre cognition conceptuelle, à savoir sa productivité et sa systémativité (Fodor 1975, 1998, 2004, Fodor et Pylyshyn

1988, Fodor et McLaughlin 1990, Davies 1991). Ainsi, si les concepts postulés par Fodor ne sont pas censés expliquer le cours de nos processus cognitifs supérieurs, par exemple la manière dont nous catégorisons, ils expliquent cependant certaines propriétés de nos attitudes propositionnelles.

Plus sérieusement, on peut adresser à Fodor *l'objection Je-choisis-ce-qui-m'arrange*. Admettons que la systématique et la productivité sont bien des propriétés de notre cognition. Il reste que Fodor choisit un phénomène particulier parmi les nombreux phénomènes qui caractérisent la cognition conceptuelle des êtres humains et suppose que les autres phénomènes (par exemple, les effets de typicalité) n'ont aucune pertinence. Fodor devrait expliquer pourquoi seuls ces deux phénomènes devraient être pertinents pour une théorie des concepts.

Fodor a défendu les idées suivantes.<sup>28</sup> Il affirme tout d'abord que la systématique et la productivité sont les seules propriétés qui sont connues à propos de nos concepts (Fodor 1998). Cette affirmation est cependant fautive. En dépit de dizaines de pages consacrées aux arguments de Fodor à propos de la productivité et de la systématique en philosophie, psychologie et intelligence artificielle (par exemple, Andler 1990, Van Gelder 1990, Smolensky 1988, 1990, 1991, Clark 1993, Récanati 2003 etc.), l'étendue et la nature de la systématique et de la productivité de notre cognition conceptuelle n'ont pas été décrites en détail et restent une question controversée. En outre, l'affirmation de Fodor selon laquelle la cognition conceptuelle est productive et systématique repose sur la systématique et la productivité apparente de la compétence linguistique humaine (Fodor et Pylyshyn 1988). Mais il n'est pas certain que le langage révèle les propriétés de notre cognition conceptuelle. Finalement, le fait que des ensembles de connaissances spécifiques occurrent dans de nombreux processus conceptuels (par exemple, la catégorisation, l'induction, la déduction, les analogies etc.) et le fait que ces ensembles de connaissances expliquent le cours de ces processus sont aussi bien établis, sinon davantage, que la systématique et la productivité de notre pensée. Ainsi Fodor ne peut-il

---

<sup>28</sup> Il se peut aussi que Fodor croit que la psychologie de nos capacités cognitives supérieures (catégorisation, induction, déduction...) est dans l'ensemble profondément erronée. En conséquence, elle ne devrait pas être la base de nos théories des concepts. Cette idée s'accorde de toute évidence avec sa thèse selon laquelle les principes de base des sciences cognitives sont inappropriés pour expliquer notre cognition centrale – et que personne ne sait comment faire (Fodor 1983, 2001). Cette vue s'accorde aussi avec son attitude méprisante à l'égard de la psychologie de la catégorisation (Fodor 2000).

pas avancer que nous ne connaissons vraiment que deux propriétés de notre cognition conceptuelle, sa productivité et sa systématique.

On pourrait proposer la justification suivante. La systématique et la productivité sont des propriétés des attitudes propositionnelles. Or les propriétés des attitudes propositionnelles résultent directement des propriétés des concepts. Car les attitudes propositionnelles sont faites de concepts. En conséquence, la systématique et la productivité sont des *explananda* plus importants pour une théorie des concepts que les données à propos de nos processus cognitifs. Cependant, cet argument repose sur une image de la cognition que je rejette (voir les sections 1.2. et 1.3. ci-dessus), à savoir l'idée que les concepts sont les constituants de nos attitudes propositionnelles. Je conclus que Fodor nous doit un argument pour justifier son rejet du principe de l'efficacité causale. En l'absence d'un tel argument, je propose d'admettre ce principe.

#### *2.4. L'hypothèse de l'espèce naturelle (l'hypothèse EN)*

La dernière hypothèse concerne la nature de la classe des représentations mentales à laquelle réfère la notion théorique de concept.

##### **L'hypothèse EN**

Les concepts forment une espèce naturelle.

Cette hypothèse exprime l'idée suivante : les concepts forment une classe homogène à propos de laquelle il est possible de formuler des généralisations inductives importantes. De toute évidence, il y a de nombreuses différences entre les concepts et entre différents types de concepts (par exemple, entre les concepts musicaux – BEMOL, ACCORD – et les concepts politiques – SOCIALISTE, ETAT, JUSTICE). Mais l'idée centrale de l'hypothèse EN est que, par delà ces différences, les concepts ont en commun de nombreuses propriétés qu'une théorie des concepts doit déterminer.

Si cette hypothèse est vraie, la notion de concept est alors importante. Il y a de nombreuses manières de classer les choses, par exemple les états mentaux. La plupart d'entre elles sont inutiles d'un point de vue scientifique. Car les classes qui en résultent ne donnent pas lieu à de nombreuses généralisations inductives. Au contraire, selon

l'hypothèse EN, la classe des représentations qui figurent par défaut dans les processus cognitifs supérieurs donne lieu à de nombreuses généralisations inductives. Donc, la notion de concept figure à juste titre dans une psychologie scientifique.

Cette idée est quelque fois explicitement admise par les psychologues.<sup>29</sup> Par exemple, Murphy écrit au début de son récent livre sur les concepts (Murphy 2002, 2-3)<sup>30</sup> :

« The psychology of concepts cannot by itself provide a full explanation of the concepts of all the different domains that psychologists are interested in. This book will not explore the psychology of concepts of persons, musical forms, numbers, physical motions, and political systems. *The details of each of these must be discovered by the specific disciplines that study them*; to fully understand people's musical concepts will require much research into the psychology of music, rather than being predictable solely from what we know of concepts per se. *Nonetheless, the general processes of concept learning and representation may well be found in each of these domains*. For example, I would be quite surprised if concepts of musical forms did not follow a prototype structure (chapter 2), did not have a preferred level of categorization (chapter 7), and did not show differences depending on expertise or knowledge (chapter 6). Spelling out what categories people have of musical forms, what levels of representations there are, and what knowledge influences the concepts is primarily part of the psychology of music rather than the psychology of concepts. But once the basic elements of musical concepts have been identified, the concepts will likely be found to follow the principles identified in other domains. In short, the psychology of concepts has much to offer other fields of psychology and cognitive science generally. »

---

<sup>29</sup> Même si elle n'est pas admise par tous les psychologues, comme nous le verrons (chapitre 2, section 3.).

<sup>30</sup> Je souligne.

De toute évidence, Murphy suppose que par delà les différences entre les types de concepts, on peut formuler de nombreuses généralisations à propos de la classe entière des concepts.<sup>31</sup>

En outre, l'hypothèse EN est implicitement acceptée par de nombreuses théories psychologiques des concepts et par la plupart des théories philosophiques. En effet, la notion de concept est une notion théorique. Elle réfère à une classe d'entités qui est postulée pour des raisons théoriques. Dans la plupart des sciences empiriques, l'introduction d'un terme théorique général présuppose que la classe dénotée est une espèce naturelle, c'est-à-dire que le terme réfère à une classe à propos de laquelle de nombreuses généralisations sont possibles. En outre, de nombreux psychologues supposent que leur théorie préférée des concepts peut expliquer la plupart des phénomènes qui caractérisent la cognition supérieure. Depuis maintenant plus de trente ans, ils s'efforcent de trouver des phénomènes explicables par leur propre théorie, mais problématiques pour les théories rivales (voir la revue de la littérature à vingt ans d'écart dans Smith et Medin 1981 et Murphy 2002). Donc, ils admettent qu'il y a une classe homogène de représentations qui sont utilisées par nos compétences cognitives supérieures.

Je ne vais pas insister pour l'instant davantage sur cette hypothèse, dans la mesure où c'est l'objet du présent travail. Je l'explique plus tard en détail (chapitre 2), puis j'examine de près ses lettres de créance (chapitres 4-8).

### *2.5. Conclusion : panorama*

Résumons ces hypothèses. Elles offrent une image cohérente de notre cognition supérieure. Il existe une classe de représentations mentales à propos de classes d'entités stockées en MLT qui figurent par défaut dans les processus cognitifs dont dépendent nos capacités cognitives supérieures. Par exemple, quand un individu doit décider si une cible est un chien, il extrait par défaut une représentation à propos des chiens. Toutes les connaissances qu'il possède à propos des chiens ne font pas partie de cette représentation. Une croyance appartient à cette représentation si elle est activée par défaut dans les

---

<sup>31</sup> Voir aussi Barsalou et al. 2003a.

processus cognitifs à propos des chiens.<sup>32</sup> En bref, il existe un ensemble de connaissances, relativement circonscrit, à propos des chiens auquel nous faisons appel par défaut dans nos performances cognitives supérieures à propos des chiens.<sup>33</sup> Ces représentations mentales, par exemple la représentation des chiens, sont utilisées par de nombreux capacités cognitives supérieures (et peut-être par la plupart d'entre elles, voire par toutes), par exemple, quand nous catégorisons des entités comme des chiens, quand nous raisonnons à propos des chiens, quand nous faisons des analogies avec les chiens etc. Les propriétés de ces représentations contribuent à expliquer causalement les processus conceptuels dans lesquels elles figurent : si Jean avait une autre représentation de chien, il ne catégoriserait pas Fido comme il le fait. Il est possible de formuler de nombreuses généralisations inductives scientifiques à propos de cette classe de représentations. Ainsi cette classe est-elle scientifiquement importante : elle doit figurer dans une psychologie scientifique. *Les concepts* sont ces représentations. L'objet d'une *théorie des concepts* est de décrire leurs propriétés générales.

---

<sup>32</sup> Pour certaines croyances, la question « est-ce que cette croyance appartient au concept de x ? » peut n'avoir qu'une réponse vague.

<sup>33</sup> Cette image de la cognition est essentiellement anti-holistique. J'ai souligné ce point plus haut.

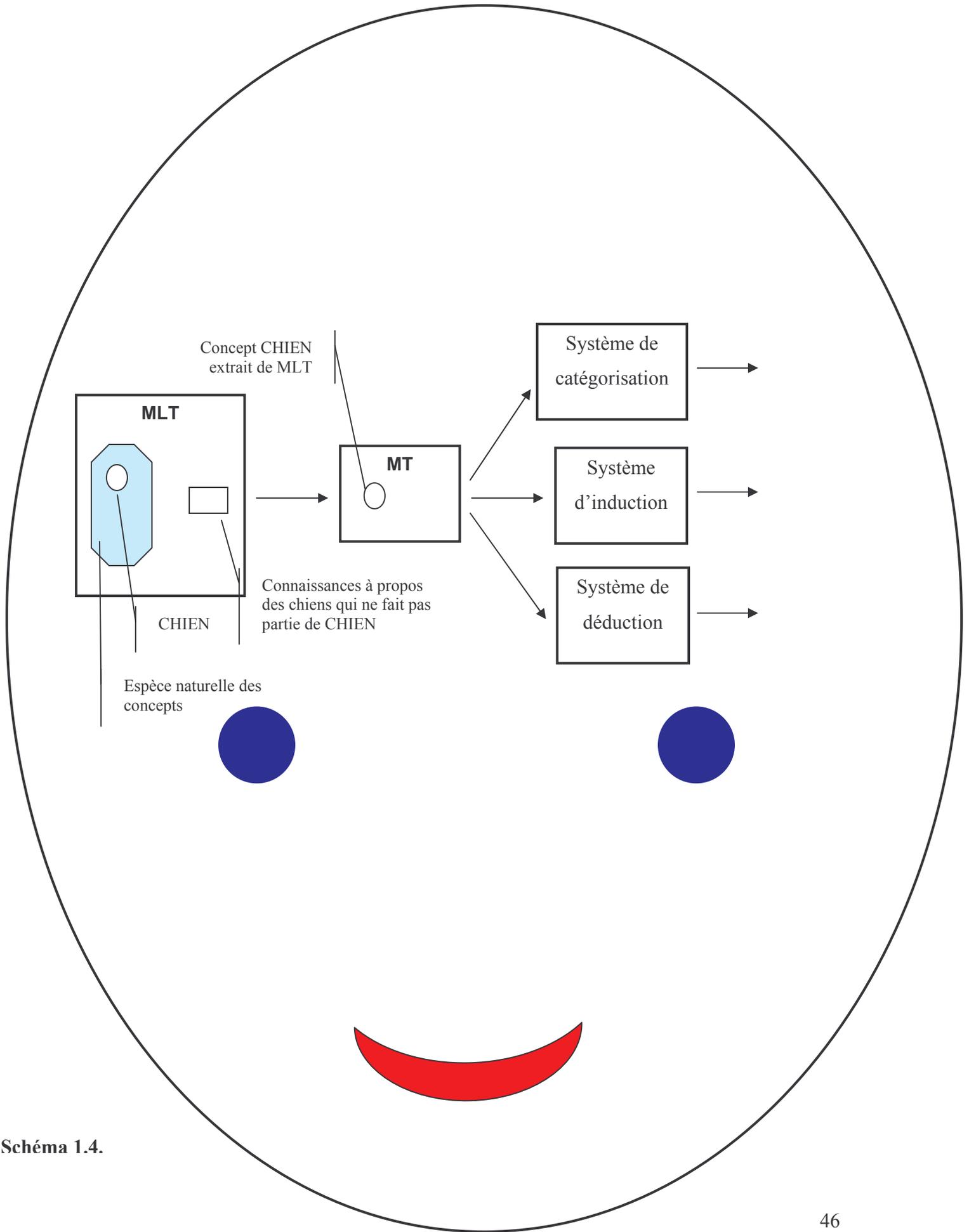


Schéma 1.4.

### **3. Qu'est-ce qu'une théorie des concepts?**

Les deux premières sections ont caractérisé la notion théorique de concept et les hypothèses à propos de la cognition qui la justifient. Dans cette section, je décris l'objet d'une théorie des concepts.

#### *3.1. Quel est l'objet d'une théorie des concepts ?*

L'hypothèse EN soutient que les concepts ont de nombreuses propriétés en commun. Elle ne nie pas l'existence de types de concepts. Mais elle suppose que par delà ces différences, il y a des propriétés communes à tous les concepts qui sont scientifiquement importantes. J'appelle ces propriétés *les propriétés générales des concepts*. Il est bien sûr important et intéressant de décrire les propriétés qui caractérisent spécifiquement différents types de concepts, par exemple nos concepts biologiques naïfs (Carey 1985, Gelman et Markman 1986, 1987, Gelman 1988, Gelman et Wellman 1991, Medin et Atran 1999 etc.). Certains psychologues des concepts sont avant tout intéressés par cette tâche. Cependant, ce n'est pas l'objet distinctif d'une théorie des concepts. Ce qui distingue une théorie des concepts de l'étude de nos concepts biologiques naïfs comme de l'étude de nos représentations visuelles ou auditives, c'est qu'elle s'efforce de déterminer les propriétés générales des concepts. La réponse à la question « Quel est l'objet d'une théorie des concepts ? » est donc la suivante.

#### **Objet d'une théorie des concepts**

Une théorie des concepts est supposée décrire les propriétés qui sont communes à tous les concepts ou à la plupart d'entre eux.

Notons qu'affirmer cela ne revient pas à soutenir que l'étude de concepts spécifiques, par exemple le concept d'objet (Spelke 1994), ou l'étude de types de concepts, par exemple les concepts biologiques naïfs, sont sans pertinence pour une théorie des concepts. Car les propriétés de concepts spécifiques ou des types de concepts doivent être compatibles avec la théorie des concepts.

Ces idées sont exprimées dans la citation de Murphy dans la section précédente. Prinz exprime également clairement cette idée (Prinz 2002, 3) :

« An adequate theory of concepts must have sufficient expressive power or breath to accommodate the large variety of concepts that we are capable of possessing. The human conceptual repertoire ranges from the sensory to the abstract. We have concepts of readily observable states within ourselves, like PAIN, theoretically derived concepts, such as ELECTRON, and seemingly formal concepts, such as NUMBER. We have concepts of natural kinds, such as FROG, artifacts, such as BOAT, and social kinds, such as MOTHER or DEMOCRACY. »

Dans ce qui suit, j'explique plus en détail les buts d'une théorie des concepts.

### 3.2. *Le contenu des concepts*

Les concepts sont des représentations mentales, c'est-à-dire des individus (des véhicules) doués de propriétés sémantiques (leur contenu). L'un des buts d'une théorie des concepts est de décrire leur contenu. Il y a plusieurs manières de s'attaquer à cette tâche. Toutes ne sont pas pertinentes pour les théories *psychologiques* des concepts.

#### 3.2.1. Quelle sorte de connaissances est encodée par les concepts ?

Un concept exprime un ensemble de propositions à propos de ce à quoi il réfère. Ce sont les connaissances qui sont extraites par défaut de la MLT pour accomplir des tâches faisant appel à nos compétences cognitives supérieures. Cela soulève la question suivante : quelles propositions sont exprimées par un concept, par exemple par CHIEN ? C'est-à-dire quelles connaissances sont extraites par défaut de la MLT quand un individu est impliqué dans une tâche cognitive à propos des chiens ? On peut poser la même question pour n'importe quel concept. Par exemple, quelles propositions sont exprimées par le concept d'une classe d'événements, aller chez le dentiste ?

La psychologie des concepts ne cherche pas à répondre à ces questions : elle ne porte pas sur les propositions exprimées par un concept particulier, par exemple par le concept CHIEN. Elle porte sur les propriétés qui caractérisent *en général* la classe des

concepts. Cependant, une question analogue peut être posée à propos de la classe des concepts en général. En effet, la plupart des théories des concepts font l'hypothèse suivante :

Tous les concepts (ou la plupart d'entre eux) expriment le même type de propositions.

Autrement dit, le type de propositions exprimé par un concept est une propriété générale des concepts.

On peut rendre cette idée un peu plus précise comme suit : les propositions exprimées par tous les concepts (ou par la plupart) ont la même *forme*. Supposons, par exemple, que les concepts expriment des définitions. Tous les concepts exprimeraient alors des propositions dont la forme serait la suivante :

[x est un  $\Gamma$  ssi  $P_1x$  et  $P_2x$  et ... et  $P_nx$ ]

Les propriétés nécessaires  $P_1$  ...  $P_n$  diffèrent pour chaque concept. Par conséquent, CHIEN et ALLER CHEZ LE DENTISTE expriment bien sûr des propositions différentes à propos respectivement des chiens et des événements consistant à aller chez le dentiste.<sup>34</sup> Mais les propositions qu'ils expriment ont la même forme.

Il est maintenant possible de caractériser l'un des buts d'une théorie des concepts. Une théorie des concepts prétend répondre à la question suivante :

*Quel type de propositions est exprimé par un concept ?*

Cette question est *empirique*. Puisque les processus cognitifs peuvent être traités comme des homonculi qui lisent le contenu des concepts et agissent en conséquence (voir le schéma 1.3.), le contenu des concepts peut être inféré, quoique de manière indirecte, à

---

<sup>34</sup> Ainsi Murphy écrit-il (Murphy 2002, 3): « (...) the general processes of concept learning and representation may well be found in each of these domains. » Je souligne.

partir du déroulement de nos processus cognitifs. Comme nous le verrons, les principales théories psychologiques des concepts offrent des réponses différentes à cette question.

D'autres questions peuvent être posées à propos des propositions qui sont exprimées par nos concepts. Plutôt que de s'intéresser à la *forme* de ces propositions, on peut se demander si elles ont un *contenu* spécifique :

*Est-ce que les concepts encodent un contenu spécifique ?*

Par exemple, les phénoménalistes supposent que nos concepts peuvent être réduits à des conjonctions, disjonctions ou encore implications de représentations de mouvements, de couleurs, de formes, de sons etc., c'est-à-dire à des représentations de propriétés perceptuelles (cette idée semble être présente dans la théorie des idées générales de Berkeley 1734 et dans la sémantique vérificationniste). Au contraire, on pourrait soutenir que les propositions qui sont représentées par nos concepts ne sont pas, ou pas uniquement, perceptuelles.

Les questions sur le contenu des propositions sont de nouveau d'actualité dans la littérature sur les concepts. Barsalou a développé une conception empiriste des concepts (voir la théorie des symboles perceptuels dans Barsalou 1999, Barsalou et al. 2003 a, b et le chapitre 3, section 3.). Il s'est efforcé de montrer que nos concepts encodent des propositions perceptuelles. Prinz a développé des idées semblables (Prinz 2002, voir le chapitre 3, section 3.).

### 3.2.2. Pourquoi les concepts possèdent-ils leurs propriétés sémantiques ?

Les concepts sont des représentations mentales qui réfèrent à quelque chose et qui encodent des connaissances à propos de ce à quoi ils réfèrent. Cela soulève la question suivante (par exemple, Fodor 1987, 1990, 1998, Millikan 1984, 2000, Prinz 2002) :

*Pourquoi les concepts possèdent-ils les propriétés sémantiques qui sont les leurs ?*

Une question analogue peut être posée à propos de n'importe quel type de représentations. Par exemple, de nombreux mots possèdent un sens et/ou une référence. Par exemple, le mot français « chien » réfère à la classe des chiens. On peut donc se demander pourquoi « chien » ou les mots du français ou, en général, les mots d'une langue naturelle, possèdent leur référence. Pour expliquer la référence de « chien », on pourrait par exemple avancer l'idée suivante :

« chien » réfère aux chiens parce que les locuteurs du français ont la même connaissance commune que « chien » est utilisé pour référer aux chiens.

Si les conventions du français étaient différentes, « chien » référerait à quelque chose d'autre, par exemple aux chevaux. Par convention, le nom français « célibataire » est synonyme avec « adulte non marié » et cette relation de synonymie définit « célibataire ». La même question peut être posée à propos des photos et des peintures : d'où viennent leurs propriétés sémantiques ?

On peut aussi poser cette question dans une perspective naturaliste. Le naturalisme est la thèse selon laquelle seules les propriétés, classes etc. qui sont posées par les sciences de la nature ou qui sont, dans un sens assez vague, compatibles avec elles, existent.<sup>35</sup> Or, de prime abord, les propriétés sémantiques ne s'accordent pas avec l'approche naturaliste (Fodor 1987). En conséquence, un naturaliste qui croit en la réalité des propriétés sémantiques doit expliquer comment des concepts peuvent avoir des propriétés sémantiques – et ce, en termes non sémantiques (Fodor 1987, Stich et Warfield 1994, Jacob 1997). Si l'on pouvait présenter des conditions suffisantes de la forme suivante :

Si un véhicule a telle et telle propriété (non sémantique), alors il réfère à x,

on aurait alors établi l'innocence naturaliste de la référence (*mutatis mutandis* pour le sens).

---

<sup>35</sup> La nature et la cohérence du naturalisme ont été largement discutées dans les années 90 (Crane et Mellor 1990, Chomsky 1995, Jackson et Braddon-Mitchell 1996, Jackson 1997).

Cette question a été au cœur de la *philosophie* des concepts. De nombreux philosophes ont admis, et admettent toujours, qu'une théorie des concepts appartient à la sémantique – c'est-à-dire à la théorie qui explique d'où viennent les propriétés sémantiques des représentations (Rey 1983, 1985, Fodor 1994). Les théories des concepts sont supposées être un élément, peut-être l'élément crucial, d'une explication de l'intentionnalité des êtres humains, c'est-à-dire de notre capacité à nous représenter comment les choses sont. Les êtres humains agissent parce qu'ils ont des états mentaux pourvus de contenus spécifiques : Jean ouvre le frigo parce qu'il veut du jus d'orange et parce qu'il croit qu'il y a du jus d'orange dans le frigo. Si l'on admet que, d'une manière ou d'une autre, les concepts sont les constituants de nos pensées, il est naturel de supposer que les propriétés sémantiques de nos pensées sont dérivées des propriétés sémantiques des concepts. Ainsi comprise, une théorie des concepts doit s'occuper principalement de deux questions :

- Quelles sont les propriétés sémantiques des concepts ?
- Pourquoi possèdent-ils ces propriétés ?

Il est surprenant de constater que de nombreux philosophes croient que les psychologues s'occupent aussi de ces questions. Autrement dit, certains philosophes pensent que les théories *psychologiques* des concepts sont des théories *sémantiques*. Par exemple, selon Fodor, les théories des prototypes avancent les thèses suivantes (Fodor 1994, 1998) :

- (i) Les concepts ont un sens et une référence.
- (ii) Le sens d'un concept est un ensemble de propriétés non nécessaires, mais typiques (voir Searle 1958 pour une sémantique de ce type).
- (iii) La référence d'un concept est l'ensemble d'entités qui possèdent un nombre suffisant de ces propriétés.

Les théories des prototypes admettent bien en effet (i) et (ii), mais n'ont pas à admettre (iii). Elles sont au contraire compatibles avec un grand nombre de théories de la référence

(Prinz 2002). Par ailleurs, les théories des prototypes n'expliquent pas comment les concepts en viennent à posséder un sens. Elles n'ont que très peu à dire sur la sémantique des concepts. C'est aussi le cas des autres théories des concepts.<sup>36</sup>

Je propose d'éviter cette erreur d'interprétation. De nombreuses données empiriques viennent à l'appui de la thèse suivante : nous avons en MLT des ensembles de connaissances, relativement bien circonscrits, que nous utilisons par défaut dans les processus cognitifs. Les psychologues les appellent des concepts. Ils s'efforcent de déterminer la nature de ces connaissances (voir la section 3.2.1.), la manière dont elles sont représentées (section 3.3.) et leur utilisation par nos processus cognitifs (section 3.5.). Mais ils ne cherchent pas à déterminer comment les concepts acquièrent leurs propriétés sémantiques.<sup>37</sup>

*Dans ce travail de thèse, j'évite autant que faire se peut les questions sémantiques.* Deux raisons justifient principalement cette position :

- Je souhaite me concentrer autant que possible sur les théories *psychologiques* des concepts et sur les questions philosophiques qu'elles posent.
- Je pense que la sémantique repose sur des fondements douteux (Stich 1996, Bishop et Stich 1998, Machery et al. 2004). Je n'offre toutefois pas dans ce travail de critique détaillée de la sémantique.

Ainsi l'orientation de ce travail sur les concepts diffère-t-elle fortement des travaux philosophiques sur les concepts. La plupart des philosophes pensent qu'une théorie des concepts devrait, au moins, inclure une explication de la manière dont les concepts possèdent leurs propriétés sémantiques. D'autres philosophes pensent que c'est là le cœur de la philosophie des concepts. Je voudrais montrer que l'approche

---

<sup>36</sup> Les psychologues des concepts peuvent bien sûr avoir des théories sémantiques en sus de leurs théories psychologiques.

<sup>37</sup> On pourrait objecter que les psychologues du développement cherchent précisément à répondre à cette question. Cela n'est cependant pas le cas. Les psychologues développementaux cherchent à comprendre comment et pourquoi les concepts des nourrissons et des enfants, par exemple le concept d'objet, se développent (Spelke 1994, Bloom 2000). Mais ils n'expliquent pas pourquoi ces concepts possèdent les propriétés sémantiques qui sont les leurs.

psychologique des concepts pose des questions intéressantes qui ont été ignorées par la philosophie traditionnelle des concepts.

### *3.3. Quelles sont les propriétés des véhicules des concepts ?*

Les concepts sont des représentations. Ils possèdent donc un véhicule qui peut être décrit de manière non sémantique (section 1.1.2.). On peut donc se demander si l'on peut formuler des généralisations concernant les véhicules des concepts. Y a-t-il des généralisations qui soient vraies de tous les véhicules des concepts ? C'est un autre objet d'une théorie psychologique des concepts. Une théorie des concepts prétend donc répondre à la question suivante :

#### *Quelles sont les propriétés générales des véhicules des concepts ?*

Cet aspect de la théorie des concepts a été développé dans plusieurs directions.

- Pendant des siècles, les philosophes se sont demandés si les êtres humains pensaient au moyen d'images, au moyen de mots ou au moyen de quelque chose d'autre.<sup>38</sup> Formulée en des termes contemporains, la question est la suivante : quand nous nous engageons dans une tâche non perceptuelle, par exemple dans la résolution d'un problème, faisons-nous appel à des représentations (peut-être propres à une modalité sensorielle particulière, « modality-specific ») qui figurent dans le traitement perceptuel des stimuli ou à un tout autre genre de représentations ?

Cette question agite les psychologues des concepts depuis l'invention de la psychologie scientifique.<sup>39</sup> Bien plus, ce débat fait toujours rage dans la psychologie contemporaine des concepts, en philosophie et entre philosophes et psychologues (Barsalou 1999, Barsalou et al. 2003a, b et Prinz 2002, voir le chapitre 3).

- L'organisation des concepts en MLT peut aussi être considérée comme faisant partie de l'étude du véhicule des concepts. La question est la suivante : comment sont

---

<sup>38</sup> Aristote, dans le *De Anima*, Locke, Berkeley et Hume ont tous soutenu que les concepts étaient des images. Guillaume d'Ockham (Panaccio 1993), puis Descartes ont rejeté cette idée.

<sup>39</sup> Machery, E., « Concept theories : Old ideas in disguise », manuscrit.

stockés les concepts en MLT ? Les psychologues ont ainsi étudié les relations d'activation entre les concepts (Collins et Loftus 1975).

- Plus récemment, l'hypothèse selon laquelle différents types de concepts (concepts biologiques, concepts d'artéfacts...) sont encodés en différentes parties du cerveau a aussi été étudiée. L'existence de déficits conceptuels spécifiques à certaines catégories est maintenant bien établie (voir la synthèse de Caramazza et Mahlon 2003 et la discussion entre Moss et Tyler 2003 et Mahlon et Caramazza 2003). Par exemple, certains sujets perdent entièrement la capacité de nommer des plantes, mais conservent la capacité de nommer des artéfacts. Toutefois, nul pattern simple et cohérent n'a encore émergé de ces dissociations. L'imagerie cérébrale peut de toute évidence éclairer cette questions (Caramazza et Mahlon 2003).

### *3.4. La méthodologie habituelle*

Selon le principe de l'efficacité causale, les propriétés des concepts contribuent à expliquer causalement les propriétés des processus conceptuels dans lesquels ils figurent. Les propriétés du concept CHIEN de Marie expliquent pourquoi elle catégorise Médor comme étant un chien. Les psychologues supposent aussi que les propriétés générales des concepts contribuent à expliquer causalement certaines propriétés qui sont généralement vraies des processus conceptuels. Par exemple, elles sont supposées expliquer causalement certaines propriétés générales de nos processus de catégorisation.

Cela suggère que l'on peut déterminer, expérimentalement, certaines propriétés des processus conceptuels et inférer de manière abductive les propriétés des concepts. La plupart des études en psychologie des concepts utilisent cette méthodologie.<sup>40</sup>

### **La méthodologie habituelle**

On peut inférer abductivement les propriétés des concepts qui figurent dans ces processus à partir des propriétés de nos processus cognitifs.

### *3.5. Les propriétés fonctionnelles des concepts*

---

<sup>40</sup> Ce n'est cependant pas la seule méthodologie. L'étude de la localisation cérébrale des concepts n'utilise pas cette méthodologie.

On ne peut inférer abductivement les propriétés des concepts à partir des propriétés des processus conceptuels que si l'on est capable de décrire comment les concepts figurent dans les processus conceptuels. Autrement dit, on doit avoir une théorie, fût-elle vague, de ces processus conceptuels. Différentes théories des concepts peuvent être compatibles avec les propriétés des processus cognitifs supérieurs si elles sont combinées avec différentes théories des processus eux-mêmes. Ainsi est-il *impossible* de développer une théorie psychologique des concepts sans développer en même temps une théorie, au moins rudimentaire, des processus conceptuels. Par conséquent, la psychologie des concepts cherche à apporter une réponse à la question suivante :

*Comment les concepts figurent-ils dans les processus conceptuels ?*

La manière dont les concepts figurent dans ces processus définit les *propriétés fonctionnelles* des concepts.

De fait, les théories des concepts ont été développées de pair avec les théories des processus conceptuels, particulièrement avec les théories de la catégorisation. Par exemple, les théories des prototypes ont remplacé les théories classiques des concepts en partie parce qu'elles étaient capables d'expliquer la corrélation entre la typicalité d'un objet et la vitesse de catégorisation de cet objet (voir le chapitre 5). Or les théories des prototypes ne peuvent à elles seules expliquer cette corrélation. Elles doivent être associées à une théorie des processus de catégorisation (comme le notent, par exemple, Laurence et Margolis 1999).

### *3.6. Autres aspects des théories psychologiques des concepts*

Bien que les trois questions présentées ci-dessus (à propos du contenu des concepts, de leurs véhicules et de leurs propriétés fonctionnelles) soient les principaux objets d'une théorie des concepts, d'autres questions sont importantes et ont été examinées par les psychologues. Cela est particulièrement le cas de la question suivante :

*Comment acquiert-on les concepts ?*

Notons que cette question est ambiguë. Elle peut concerner l'acquisition des concepts par les enfants ou l'acquisition des concepts par les adultes. Les deux réponses ne sont pas forcément identiques.

### *3.7. Conclusion*

Les théories *psychologiques* des concepts ne cherchent pas à répondre à toutes les questions à propos des concepts. En particulier, certaines questions sémantiques sont proprement philosophiques. Les théories psychologiques des concepts s'occupent donc principalement de trois questions :

*Quels types de propositions sont exprimés par les concepts ?*

*Quelles sont les propriétés générales des véhicules des concepts ?*

*Comment les concepts figurent-ils dans les processus cognitifs supérieurs ?*

## **4. Conclusion**

La psychologie des concepts admet une image particulière de l'esprit humain. Les concepts sont des ensembles de connaissances à propos d'entités spécifiques. Ces ensembles de connaissances sont utilisés par un grand nombre de capacités cognitives supérieures : nous faisons appel à eux quand nous catégorisons, déduisons, induisons, faisons des analogies etc. Il est vrai que nous pouvons utiliser n'importe quel type de connaissances dans nos processus cognitifs supérieurs. C'est la part de vérité du holisme des processus cognitifs. Mais la psychologie des concepts soutient qu'il existe des ensembles de connaissances relativement bien définis qui sont utilisés par défaut dans nos processus. Un concept de *x* contient les connaissances qui sont utilisées par défaut dans les processus cognitifs à propos de *x*. Ces ensembles de connaissances ont en commun un grand nombre de propriétés : ils constituent une classe pertinente pour une psychologie scientifique, c'est-à-dire ils forment une espèce naturelle. La psychologie des concepts cherche à déterminer les propriétés générales de ces ensembles de connaissances. Elles portent particulièrement sur le contenu de ces représentations, sur leurs véhicules et sur la manière dont ils figurent dans les processus cognitifs supérieurs.

La correction de cette image est de toute évidence une question empirique. Est-ce que notre connaissance empirique à propos des processus cognitifs supérieurs est compatible avec elle ? Si nous devons rejeter ou modifier cette image, quels aspects devons-nous conserver et lesquels devons-nous rejeter ? Dans ce travail, j'apporte une réponse partiellement négative à la première question.

Il devrait finalement être évident qu'une discussion philosophique de la notion scientifique de concept est indissociable d'une discussion empirique de cette image. En psychologie comme ailleurs, la philosophie des sciences ne peut être séparée des questions empiriques. Cette approche des concepts s'écarte délibérément de la plupart des discussions philosophiques des concepts. Je n'aborderai ni la question de l'individuation des concepts ni celle de leur sémantique. Je crois que la validité de la notion de concept pour une psychologie scientifique est une question aussi intéressante, si ce n'est plus, que les questions posées habituellement en philosophie. J'espère que ce travail justifiera cette affirmation !

## Chapitre 2

### Les concepts ne forment pas une espèce naturelle

---

Le premier chapitre peint l'arrière-plan de la thèse que je défends dans le présent travail. J'ai expliqué la notion de concept et les buts des théories des concepts. Dans ce chapitre, je présente en détail la thèse que j'entends défendre. Les théories des concepts cherchent à *caractériser les propriétés générales des concepts*, c'est-à-dire les propriétés qui sont communes à tous les membres de la classe des concepts (ou au moins à la plupart d'entre eux). Ce faisant, elles supposent que la classe des concepts forme *une classe homogène de représentations mentales*. Cela explique pourquoi la classe des concepts est traitée comme *une classe théoriquement importante*. Ces idées sont exprimées par *l'hypothèse EN* :

#### **L'hypothèse EN**

Les concepts forment une espèce naturelle.

Bien que la plupart des philosophes et de nombreux psychologues acceptent, implicitement ou explicitement, cette hypothèse, ce travail se propose de la critiquer. En effet, je prétends défendre la thèse suivante :

#### **Thèse principale**

*Des données empiriques en nombre croissant suggèrent que les concepts ne forment pas une espèce naturelle.*

Contrairement à ce que l'on a admis, il n'y a pas de classe homogène de représentations mentales dont les propriétés communes expliqueraient de manière causale les propriétés des processus cognitifs supérieurs. Au contraire, les représentations qui figurent dans les processus cognitifs supérieurs se divisent en types qui ont peu en commun.

L'objet de ce chapitre est de détailler cette position. Dans la première section, j'explique en détail l'hypothèse EN. J'examine quelles propriétés une classe doit posséder pour être une espèce naturelle. Dans la deuxième section, j'explique sur cette base pourquoi je tiens le rejet de l'hypothèse EN pour justifier. Pour ce faire, je présente une stratégie empirique qui cherche à montrer que la classe des concepts ne possède pas les propriétés requises pour être une espèce naturelle. Finalement, je compare ma proposition avec des idées semblables qui ont été avancées par plusieurs psychologues.

## **1. Qu'est-ce qu'une espèce naturelle ?**

Dans cette section, j'analyse la notion d'espèce naturelle : que veut-on dire quand l'on qualifie une classe d'espèce naturelle ? Je cherche une notion d'espèce naturelle qui soit pertinente pour caractériser l'hypothèse EN : en particulier, cette notion ne doit pas être trop restrictive. Autrement, le rejet de l'hypothèse EN serait trivial.

La terme « espèce naturelle » a été utilisé dans trois disciplines, en psychologie, en sémantique et en philosophie des sciences. Dans ces trois disciplines, il a différents sens. Examinons-les une par une.

### *1.1. La notion psychologique d'espèce naturelle*

On trouve en psychologie une première notion d'espèce naturelle (par exemple, Keil 1989, Gelman et Coley 1991, 150-158, Gopnik et Meltzoff 1997, chap. 6). Dans cette discipline, les espèces naturelles sont par stipulation les classes d'objets qui ne sont pas des artefacts, particulièrement les classes de plantes, d'animaux et de matières, et qui sont dénotées par des noms communs. Les espèces naturelles sont par stipulation opposées aux classes d'artefacts (par exemple, les chaises) et non aux classes nominales (par exemple, la classe des objets physiques rouges qui pèsent plus de deux kilos).

- (1) Une classe est une espèce naturelle si et seulement si elle est dénotée par un nom commun et est constituée d'entités qui ne sont pas des artefacts.

La notion d'espèce naturelle, ainsi comprise, est importante en psychologie parce que les espèces naturelles sont supposées être l'objet d'une espèce particulière de

connaissance naïve, différente de notre connaissance naïve à propos d'autres classes, en premier lieu les artéfacts. Par exemple, certains psychologues affirment que les humains croient que les espèces naturelles, mais non les artéfacts, possèdent une essence (voir Atran 1990, Gelman et Hirschfeld 1999, voir aussi le chapitre 7). Autrement dit, la notion d'espèce naturelle en psychologie suppose que l'on puisse distinguer différents types de connaissance : ce sont les propriétés de l'esprit et non des classes elles-mêmes qui motivent l'introduction de cette notion.

### *1.2. La notion sémantique d'espèce naturelle*

La notion d'espèce naturelle est aussi utilisée en sémantique (par exemple, Schwartz 1980). Une classe est une espèce naturelle si et seulement si elle est dénotée par un terme d'espèce naturelle. Sinon, c'est une classe nominale. Les termes d'espèce naturelle et les termes de classe nominale sont distingués de manière sémantique. C'est-à-dire, ils ont des propriétés sémantiques différentes.

- (2) Une classe est une espèce naturelle si et seulement si elle est dénotée par un terme d'espèce naturelle, c'est-à-dire par un terme avec une sémantique distinctive (par opposition aux termes de classe nominale).

Par exemple, Schwartz (1978) a suggéré la distinction suivante. Une expression linguistique est un terme d'espèce naturelle si et seulement si elle peut figurer dans des énoncés pour lesquels on a l'intuition qu'ils sont nécessaires s'ils sont vrais, bien qu'ils soient épistémiquement contingents, donc corrigibles. « L'eau est H<sub>2</sub>O » et « les humains sont des animaux » sont des énoncés de ce type : on a l'intuition qu'ils sont nécessaires s'ils sont vrais. Cependant, ils pourraient se révéler faux : ils sont épistémiquement contingents. Si une expression linguistique ne peut figurer dans des énoncés de ce type, c'est un terme de classe nominale. Les énoncés nécessaires qui impliquent des termes d'espèce nominale sont analytiques, donc incorrigibles : une proposition qui attribue une propriété à une classe nominale est analytiquement nécessaire ou contingente. Selon Schwartz, nos intuitions modales déterminent si des expressions linguistiques sont des

termes d'espèce naturelle ou des termes de classe nominale. Donc, pour Schwartz, nos intuitions modales déterminent si une classe est nominale ou naturelle.

En sémantique, la notion d'espèce naturelle suppose donc que l'on puisse distinguer des types de mots en fonction de leurs propriétés sémantiques : ce sont les propriétés des mots, et non des classes elles-mêmes qui motivent son introduction.

Les notions psychologique et sémantique d'espèce naturelle sont sans aucun doute utiles en psychologie et en sémantique. Mais elles ne sont guère utiles pour mon propos. Car je ne m'intéresse ni à la sémantique des expressions linguistiques ni aux différents types de connaissance naïve : je m'intéresse ici aux classes qui sont dénotées par les prédicats scientifiques.

### *1.3. La notion philosophique d'espèce naturelle*

#### 1.3.1. L'idée principale

On trouve *en philosophie des sciences* une autre notion d'espèce naturelle.<sup>1</sup> L'objet de cette notion est de mettre en valeur les propriétés spéciales des classes qui sont dénotées par les termes scientifiques.<sup>2</sup> Dire d'une classe que c'est une espèce naturelle, c'est affirmer qu'elle possède les propriétés qui la rendent pertinente pour une certaine science. *Nier qu'une classe soit une espèce naturelle, c'est nier qu'elle possède les propriétés qui la rendent scientifiquement pertinente.*

La notion d'espèce naturelle est en effet intrinsèquement liée au problème de l'induction (Quine 1977, Hacking 1991). Le problème de l'induction peut être posé abstraitement comme suit :

Comment peut-on justifier l'hypothèse qu'une propriété P qui est possédée par les membres d'une sous-classe S est aussi possédée par tous les membres de C, S étant incluse dans C (ou par la plupart des membres de C) ?

---

<sup>1</sup> Hacking (1991) offre une recension partielle, mais stimulante de l'histoire de cette notion. Norton (2001) offre une recension critique des principales théories de l'induction.

<sup>2</sup> Tout le monde n'admet pas que la notion d'espèce naturelle soit centrale en philosophie des sciences. Par exemple, Russell (1948) a soutenu que cette notion est inutile pour caractériser les sciences développées, par exemple la physique. De fait, en physique, les grandeurs semblent plus importantes que les classes.

Par exemple, comment puis-je justifier l'hypothèse (fausse) que tous les cygnes (ou la plupart d'entre eux) sont blancs, étant donné que tous les cygnes que j'ai vus sont de cette couleur ?

L'existence d'espèces naturelles apporte une solution partielle au problème de l'induction. Autrement dit, elle explique en partie pourquoi je suis justifié à supposer que les entités que je n'ai pas observées possèdent les propriétés qui caractérisent les entités observées. L'idée est la suivante : certaines classes, les espèces naturelles, sont telles que leurs membres ont de nombreuses propriétés en commun, tandis que les membres d'autres classes, par exemple la classe des objets qui pèsent plus de trente kilos, n'ont que très peu de propriétés en commun. Ainsi, si les cygnes forment une espèce naturelle, ils ont de nombreuses propriétés en commun. En conséquence, si *être blanc* est le type de propriétés que les cygnes ont en commun, il est légitime de supposer que tous les cygnes (ou la plupart d'entre eux) sont blancs.<sup>3</sup> De même, le fait que l'eau soit une espèce naturelle explique pourquoi « H<sub>2</sub>O » peut être utilisé dans de nombreux énoncés nomologiques, par exemple dans l'équation « HCl + H<sub>2</sub>O → H<sub>3</sub>O<sup>+</sup> + Cl<sup>-</sup> ».<sup>4</sup>

Par conséquent, la notion d'espèce naturelle est utilisée pour dénoter les classes à propos desquelles de nombreuses généralisations inductives peuvent être formulées. Ce sont les classes pertinentes en science (ou, au moins, dans de nombreuses sciences). Les sciences sont supposées identifier les classes dont les membres ont en commun i/ de manière non accidentelle ii/ un grand nombre de propriétés ou de relations<sup>5</sup> en sus des propriétés ou relations qui sont utilisées pour les identifier (Mill 1905).<sup>6</sup> On peut donc proposer la condition nécessaire suivante :

---

<sup>3</sup> Deux problèmes ne sont pas résolus par l'existence d'espèces naturelles :

- Comment ma croyance que C est une espèce naturelle est-elle justifiée ?
- Comment ma croyance que P est le type de propriétés qui sont vraies de tous ou de la plupart des membres de C est-elle justifiée (Goodman 1955, Quine 1977, Boyd 1991) ?

Je crois que l'induction de second ordre proposée par Quine offre une réponse satisfaisante à la première question et que la notion d'« enracinement » (« entrenchment ») de Goodman offre une réponse satisfaisante à la seconde. Je ne défends pas ces thèses ici. Pour d'autres idées sur l'induction, voir Norton 2001.

<sup>4</sup> Les matières, comme l'eau ou l'or, peuvent être traitées comme des classes (Quine 1960, 90 sq.).

<sup>5</sup> Qui ne sont pas logiquement reliées les unes aux autres.

<sup>6</sup> De toute évidence, cette explication ne serait complète que si je clarifiais la notion de propriété. Car deux individus quelconques ont en commun un nombre infini de propriétés : deux humains pèsent moins d'une tonne, moins de deux etc. Pour résoudre ce problème, on pourrait faire appel à la notion de propriété réelle

- (3) Une classe C est une espèce naturelle seulement si les membres de C ont en commun, de manière non accidentelle, un grand nombre de propriétés – en sus des propriétés qui sont utilisées pour identifier C.

Le contraste est flagrant entre les espèces naturelles et les classes nominales, dont les membres ont en commun seulement les propriétés qui sont utilisées pour identifier ces classes.<sup>7</sup> Par exemple, le terme aristotélicien « sublunaire » dénote une classe nominale, puisque les objets qui sont au-delà de la lune n'ont pas de propriétés en commun sinon *être au-delà de la lune*. C'est une classe sans intérêt en science, puisqu'elle ne peut pas être utilisée dans des généralisations scientifiques pertinentes.

Cette idée a été développée de plusieurs manières. Ces développements ne s'entendent pas sur ce qui rend non accidentelle la possession commune de nombreuses propriétés par les membres des espèces naturelles. Puisque l'élément central dans l'argument que je développe est l'idée que les membres des espèces naturelles ont de nombreuses propriétés en commun, il n'est pas nécessaire de s'étendre longuement sur ces différentes notions d'espèce naturelle. Cependant, pour écarter certaines incompréhensions, un bref résumé de ces différentes notions est utile. En particulier, une espèce naturelle n'a ni à posséder une essence ni n'a à donner lieu à des lois *stricto sensu*.

### 1.3.2. Espèce naturelle et essence

La première notion associe les espèces naturelles aux essences.

- (4) Une classe d'entités forme une espèce naturelle seulement si les membres de C ont en commun un grand nombre de propriétés parce qu'elles ont une essence commune.

---

(Lewis 1983) ou à une solution goodmanienne (Goodman 1955) : les propriétés pertinentes sont celles qui sont dénotées par des prédicats fréquemment utilisés (« entrenched », Hacking 1991, note 8, 124-125).

<sup>7</sup> Il faut entendre ici comme ailleurs « propriétés pertinentes ». Par exemple, des modifications triviales des propriétés utilisées pour identifier les classes nominales, par exemple des propriétés disjonctives, ne sont pas pertinentes.

L'essence d'une classe est un ensemble de propriétés intrinsèques et causalement explicatives dont la possession est nécessaire et conjointement suffisante pour appartenir à cette classe (Locke 1690/1964, 286<sup>8</sup>, Kripke 1972, Putnam 1975, pour une critique de cette notion, voir Mellor 1977). Les éléments, les minéraux et les matériaux sont des espèces naturelles en ce sens : x est un morceau d'or si et seulement s'il est fait d'atomes dont le nombre atomique est 79. Cette constitution matérielle explique les autres propriétés de l'or (par exemple, le fait qu'il est oxydé par le mercure et par l'eau). De manière générale, de nombreuses notions en physique et en chimie dénotent des espèces naturelles en ce sens (par exemple, « électron », « gaz »...).

### 1.3.3. Espèce naturelle et loi

La seconde notion associe les espèces naturelles et les lois.

- (5) Une classe d'entités forme une espèce naturelle seulement si elle donne lieu à des lois *stricto sensu*.

La notion de loi doit être prise ici au sens strict. Une loi est un énoncé général qui décrit des relations entre événements, états ou grandeurs, qui n'est pas *ceteris paribus*, qui n'est pas temporellement et spatialement restreint et qui implique des propositions contrefactuelles (autrement dit, elle ne décrit pas simplement la réalité actuelle). Cette notion d'espèce naturelle s'applique peut-être aux particules fondamentales qui sont dénotées en physique (Collier 1996).<sup>9</sup>

Ces deux notions distinguent les espèces naturelles des classes nominales, puisque les membres des classes nominales ont peu de propriétés en commun. Elles distinguent aussi les espèces naturelles d'autres classes dont les membres ont des propriétés en commun par accident. Enfin, elles les distinguent de certaines classes dont les membres

---

<sup>8</sup> Locke écrit : « By this real essence, I mean that real constitution of anything, which is the foundation of all those properties that are combined in, and are constantly found to coexist with, the nominal essence : that particular constitution which everything has within itself, without any relation to anything without it. »

<sup>9</sup> Les notions d'espèce naturelle caractérisées par (4) et (5) ne sont pas équivalentes, car une classe peut donner lieu à des lois qui sont *indépendantes* les unes des autres. Ces propriétés ne sauraient alors pas les effets d'une essence.

ont des propriétés en commun de manière non accidentelle, par exemple des artefacts. Les voitures ont en commun, de manière non accidentelle, des propriétés. Mais cette corrélation n'est pas l'effet de propriétés intrinsèques. Elle résulte de causes fonctionnelles et sociologiques. En outre, la classe des voitures ne donne pas lieu à des lois au sens strict.

Ces notions d'espèce naturelle sont inadéquates dans le présent contexte. Tout d'abord, elles ne sont apparemment applicables qu'en physique et en chimie. Par exemple, elles ne peuvent s'appliquer largement en biologie. En particulier, les espèces animales et végétales ne sont pas des espèces naturelles en ce sens (Hull 1978, Griffith 1997, chap. 7-8, Sober 1999, chap. 6). Ensuite, elles ne sont pas pertinentes pour caractériser des classes d'entités psychologiques. Car il est largement admis que les classes d'entités psychologiques sont fonctionnelles (Putnam 1967). Par conséquent, elles n'ont pas d'essence. Puisque les concepts sont des entités psychologiques, la notion d'espèce naturelle caractérisée par (4) ne peut s'appliquer aux concepts. Il est aussi largement admis que, comme les autres sciences spéciales, la psychologie ne cherche pas à formuler des lois, mais des généralisations *ceteris paribus* (Fodor 1974). Puisque la notion de concept est psychologique, elle n'est pas utilisée pour formuler des lois au sens strict. Donc, la notion d'espèce naturelle caractérisée par (5) ne s'applique pas aux concepts. Si une classe devait posséder une essence ou donner lieu à des lois pour être une espèce naturelle, trivialement, « concept » ne dénoterait pas une espèce naturelle. La négation de l'hypothèse EN serait sans intérêt.

#### 1.3.4. La notion causale d'espèce naturelle

L'idée principale de la notion d'espèce naturelle en philosophie des sciences (voir (3)) peut être développée sans faire appel aux notions d'essence ou de loi. Les espèces naturelles peuvent être caractérisées comme suit (Boyd 1990, 1991, Griffith 1997) :

- (6) Une classe d'entités est une espèce naturelle si et seulement si c'est la classe maximale d'entités qui ont en commun ou qui tendent à avoir en commun un ensemble de propriétés à cause d'un ou plusieurs mécanismes causaux.

Expliquons cette définition. Une espèce naturelle est une classe à propos de laquelle des généralisations peuvent être faites : ses membres tendent à avoir des propriétés en commun.<sup>10</sup> Ces généralisations ne sont pas accidentelles : il y a au moins un mécanisme causal qui explique pourquoi ses membres tendent à avoir ces propriétés en commun. Enfin, cette classe n'est pas un sous-ensemble d'une classe à propos de laquelle on pourrait faire les mêmes généralisations.

Trois points sont importants. Tout d'abord, les propriétés qui caractérisent une espèce naturelle n'ont pas à être possédées par *tous* les membres de cette classe. Il est seulement requis que les membres d'une espèce naturelle *tendent* à avoir ces propriétés. Ainsi, les termes d'espèce naturelle n'ont pas à figurer dans des lois au sens strict. Il est seulement requis qu'elles figurent dans des généralisations *ceteris paribus*. Ensuite, les essences sont seulement l'un des mécanismes causaux par lesquels les membres d'une espèce naturelle ont en commun ou tendent à avoir en commun de nombreuses propriétés. Boyd a suggéré que l'homéostasie est un autre mécanisme (par exemple, Boyd 1990, 1991) : dans certaines espèces naturelles, l'occurrence d'une propriété cause l'occurrence d'autres propriétés et est causée par leur occurrence. En conséquence, ces propriétés tendent à être associées. Griffith a suggéré d'autres mécanismes (Griffith 1997). La descendance commune semble être la base causale de l'ensemble de propriétés qui caractérisent les espèces (caractérisées cladistiquement). Les membres d'une espèce donnée ont en commun des propriétés parce qu'ils descendent de la même population. Les causes sociologiques peuvent aussi être des mécanismes causaux. Par conséquent, les artefacts sont aussi des espèces naturelles : en vertu, entre autres, de certaines causes sociologiques, les artefacts ont en commun de nombreuses propriétés.<sup>11</sup>

La notion causale d'espèce naturelle a une large extension : les matières, par exemple, l'or, les particules physiques, par exemple, les atomes, les espèces, par exemple, *macaca fuscata*, les clades en général, par exemple, les mammifères, et les

---

<sup>10</sup> Cette notion peut être interprétée de manière probabiliste.

<sup>11</sup> Pour distinguer ces différents types d'espèce naturelle, Griffith suggère que les espèces naturelles forment un continuum entre les espèces naturelles essentialisées et les classes nominales. Bien qu'elles soient d'authentiques espèces naturelles, les classes d'artefacts sont moins naturelles que les espèces biologiques. Malheureusement, Griffith ne développe pas la question des critères de naturalité.

artéfacts sont des espèces naturelles. Cependant, cette notion n'est pas vide. Tout d'abord, on peut introduire des degrés de naturalité pour distinguer ces espèces naturelles. Ensuite, les classes nominales, par exemple la classe des objets physiques qui pèsent plus de trente kilos, ne sont pas des espèces naturelles selon cette définition, puisque leurs membres n'ont pas en commun de nombreuses propriétés en sus des propriétés utilisées pour les identifier. Pareillement, de nombreux sous-ensembles d'espèces naturelles, par exemple les chiens blancs, ne sont pas des espèces naturelles. En effet, les chiens blancs ne possèdent pas des propriétés spécifiques par rapport à l'espèce naturelle que forment les chiens – si ce n'est, bien sûr, la propriété *être blanc* qui nous sert à les identifier. Les propriétés qui tendent à être vraies des chiens blancs sont les propriétés qui tendent à être vraies des chiens : la classe des chiens blancs n'est donc pas la classe *maximale* des entités qui ont ces propriétés.

Contrairement aux autres notions d'espèce naturelle, la notion causale est pertinente pour mon propos. En effet, les classes d'entités psychologiques peuvent être des espèces naturelles selon cette notion : les classes fonctionnelles ne sont pas exclues et les espèces naturelles n'ont pas à donner lieu à des lois au sens strict. En outre, l'applicabilité de cette notion n'est pas restreinte : toute classe qui donne lieu à des généralisations non accidentelles est naturelle. *Dans ce qui suit, j'utilise la notion d'espèce naturelle exclusivement en ce sens.*

#### *1.4. Conclusion*

Dans cette section, j'ai caractérisé la notion d'espèce naturelle. Les espèces naturelles sont les classes qui donnent lieu à des généralisations inductives : on peut projeter une propriété depuis un sous-ensemble vers la classe entière, parce que les membres de cette classe ont de nombreuses propriétés en commun. Cela explique pourquoi les espèces naturelles sont dénotées par des termes scientifiques. J'ai expliqué en quoi les espèces naturelles n'ont pas à posséder d'essence ni à donner lieu à des lois au sens strict. Leurs membres doivent seulement tendre à posséder de nombreuses propriétés distinctives en commun en raison d'un mécanisme causal.

## **2. Pourquoi les concepts ne forment-ils pas une espèce naturelle ?**

Dans cette section, j'applique l'analyse conceptuelle de la notion d'espèce naturelle à l'hypothèse EN.

### *2.1. Que signifie l'hypothèse EN ?*

Soutenir que les concepts forment une espèce naturelle, c'est affirmer que la classe des concepts possède les propriétés qui caractérisent les classes pertinentes en science. Quelles sont ces propriétés ? La notion causale d'espèce naturelle implique que les notions d'essence et de loi ne sont pas pertinentes. Au contraire, elle implique que la classe des concepts possède les trois propriétés suivantes :

- (i) Il y a un large ensemble de propriétés S que les représentations mentales qui figurent dans nos processus cognitifs supérieurs tendent à posséder.
- (ii) Cela résulte d'un ou plusieurs mécanismes causaux.
- (iii) Cette classe de représentations n'est pas strictement incluse dans une classe à propos de laquelle (à peu près) les mêmes généralisations peuvent être formulées.

Il est important de souligner que les psychologues qui acceptent, implicitement ou explicitement, l'hypothèse EN n'ont pas à nier l'existence de différences entre concepts ou entre types de concepts. De toute évidence, il y a de nombreuses différences entre les concepts et entre différents types de concepts : par exemple, les concepts des artefacts ont des propriétés particulières qui ne sont pas partagées par les concepts d'animaux.<sup>12</sup> Néanmoins, en sus de ces différences, un grand nombre de propriétés sont possédées par tous les concepts ou la plupart d'entre eux. Si ces propriétés ne sont pas partagées accidentellement, les concepts forment alors une espèce naturelle.

### *2.2. Comment établir que les concepts ne forment pas une espèce naturelle ?*

Nier que les concepts forment une espèce naturelle, c'est affirmer que l'une des trois possibilités suivantes est réalisée :

---

<sup>12</sup> Par exemple, Murphy accepte explicitement l'hypothèse EN tout en soutenant qu'il y a des différences intéressantes entre types de concepts (Murphy 2002, 2-3). C'est une position parfaitement cohérente.

- (i) Les concepts n'ont pas en commun un grand nombre de propriétés.
- (ii) Ils ont bien des propriétés en commun, mais cela est accidentel.
- (iii) Ils ont bien des propriétés en commun, mais la classe des représentations mentales utilisées par défaut dans nos processus cognitifs supérieurs, c'est-à-dire les concepts, est incluse dans une classe plus vaste de représentations qui possèdent les mêmes propriétés ; autrement dit, aucun ensemble de propriétés ne distingue la classe des concepts.

Dans le présent travail, je défends la première option : *je soutiens que les concepts n'ont pas en commun un grand nombre de propriétés.*

### *2.3. Pourquoi les concepts n'ont-ils pas de nombreuses propriétés en commun ?*

Je soutiens que les concepts ne forment pas une espèce naturelle parce qu'ils n'ont pas de nombreuses propriétés en commun. Mais est-ce vrai ? Pourquoi est-ce le cas ? *Parce que la classe des concepts se divise en plusieurs types qui ont très peu en commun.* Il existe plusieurs types de concepts, et ces types de concepts ont peu en commun. J'appelle cette thèse *l'hypothèse de l'hétérogénéité (HH)* :

- (HH) La classe des concepts se divise en plusieurs types de représentations qui ont peu en commun.

L'hypothèse de l'hétérogénéité est cependant encore insuffisamment précise. J'examine trois interprétations possibles de cette hypothèse.

#### *a. Différentes capacités cognitives supérieures reposent sur des types différents de concepts*

Par exemple, la catégorisation pourrait requérir un type de concepts, tandis que le raisonnement déductif pourrait en requérir un autre et le raisonnement inductif un troisième etc. Ces types de concepts auraient peu en commun. Le schéma suivant exprime cette idée :

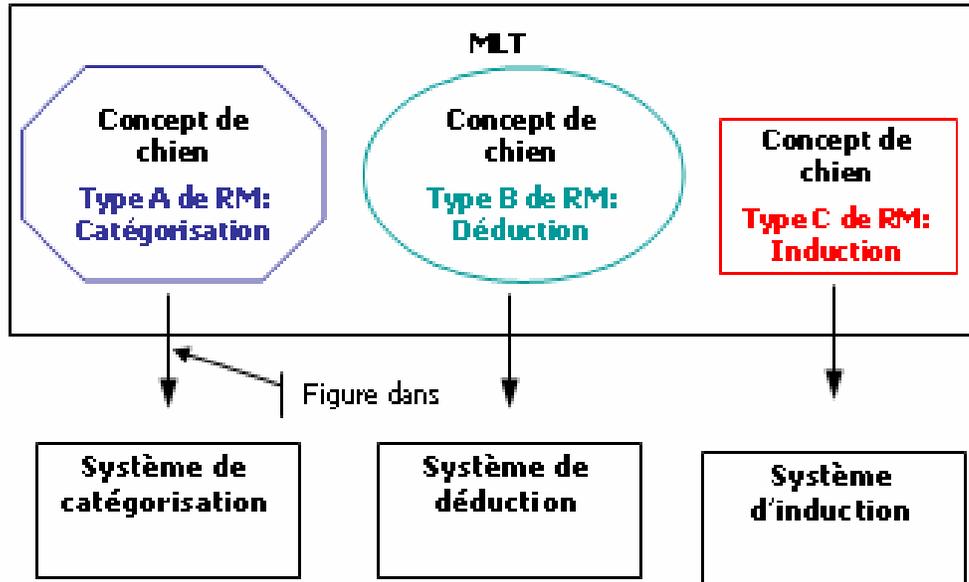


Schéma 2.1.

Je rejette cette interprétation de l'hypothèse de l'hétérogénéité. En effet, de nombreuses données empiriques suggèrent que les concepts sont utilisés par de nombreuses capacités cognitives (voir le principe de promiscuité dans le premier chapitre, section 2.2.).

*b. Différents types de catégories peuvent être représentés par différents types de concepts.*

On peut distinguer différents types de catégories de plusieurs manières.<sup>13</sup> Par exemple, on peut distinguer les catégories d'artéfacts des catégories d'animaux. D'autres classifications sont bien sûr possibles.

<sup>13</sup> « catégorie » est le nom utilisé par les psychologues pour parler des classes d'objets auxquelles réfèrent les concepts.

Différents types de catégories (par exemple, les catégories d'artéfacts, les catégories sociales, les catégories d'animaux, les catégories de plantes etc.) pourraient être représentés de différentes manières. Chaque type de catégories (par exemple, les catégories sociales et les catégories d'artéfacts) pourrait être représenté par un type homogène de concepts (les concepts des catégories d'artéfacts et les concepts des catégories sociales). Ces différents types de concepts auraient peu en commun.<sup>14</sup> Cette idée correspond au schéma suivant :

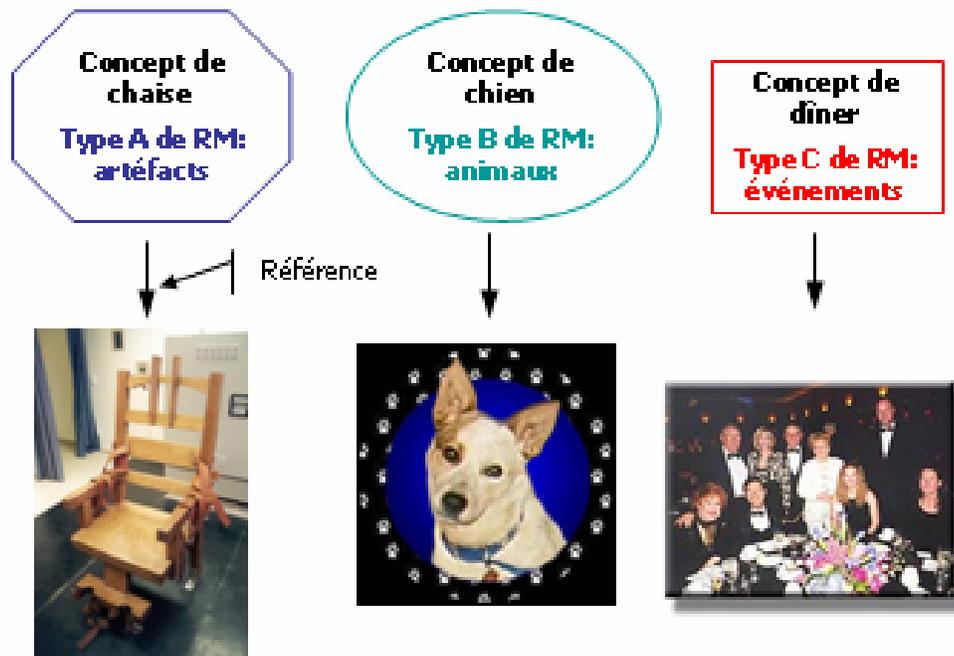


Schéma 2.2.

Certaines données empiriques suggèrent en effet que différents types de catégories sont représentés par différents types de concepts. Particulièrement, les concepts d'artéfacts et les concepts d'animaux diffèrent. Les concepts superordonnés et les concepts du niveau de base diffèrent également (Keil 1989, 1991, Ross et Makin 1999, 225-228).<sup>15</sup>

<sup>14</sup> D'autres schèmes classificatoires sont possibles (Millikan 1998, 2000).

<sup>15</sup> Les psychologues distinguent les concepts du niveau de base, par exemple CHIEN, CHAT, des concepts superordonnés, par exemple ANIMAL ou VEHICULE. Il existe plusieurs manières de définir ces niveaux (Murphy et Lassaline 1997).

Néanmoins, je n'insiste pas sur cette idée dans le présent travail. Car il n'est pas sûr que ces différences soient telles que les concepts se divisent en types qui ont vraiment peu en commun.

c. *La plupart des catégories sont représentées par de nombreux concepts qui appartiennent à des types différents.*

C'est la version de l'hypothèse de l'hétérogénéité que je défends dans le présent travail. De nombreuses catégories sont dénotées par plusieurs concepts. Et ces concepts appartiennent à des types de représentations qui ont peu en commun les uns avec les autres. Par exemple, il y a plusieurs concepts de chiens qui coréfèrent à la classe des chiens. Ces concepts de chien appartiennent à des types de représentations qui ont peu en commun. Le schéma suivant résume cette idée.<sup>16</sup>

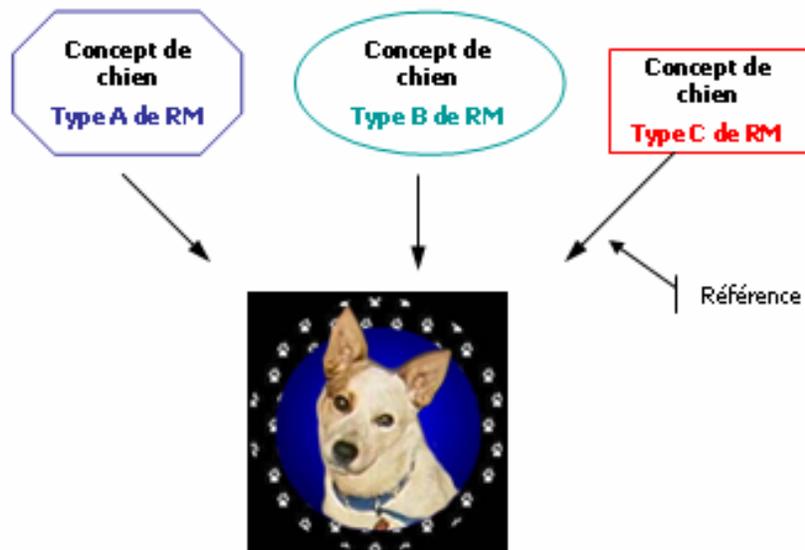


Schéma 2.3.

Ce schéma s'oppose au schéma 1.2. du chapitre 1. Cette troisième interprétation de l'HH n'est pas identique à l'idée familière, proposée par Frege, selon laquelle un

<sup>16</sup> J'explique dans le chapitre suivant quels sont, selon moi, ces types de concepts. Pour anticiper, je soutiens que ces trois types de concepts sont des prototypes, des exemplaires et des théories.

réfèrent peut être dénoté par plusieurs concepts. Car la thèse soutenue ici est que les catégories sont typiquement dénotées par des concepts *qui appartiennent à des types de représentations qui ont peu en commun les uns avec les autres*.

#### 2.4. *L'argument de la nécessité explicative (ANE)*

J'ai expliqué en détail la position que je défends dans cette thèse. Je présente maintenant la stratégie que je vais mettre en œuvre pour la défendre. L'idée principale est la suivante : il est nécessaire de postuler l'existence de plusieurs types de représentations qui ont peu en commun les uns avec les autres *si l'on veut expliquer nos compétences cognitives supérieures*. En conséquence, l'image traditionnelle d'une classe homogène de représentations gouvernant les processus conceptuels est erronée. Si « concept » est le nom de cette classe, ces concepts ne forment pas une espèce naturelle. Je présente cette idée comme suit :

##### **L'argument de la nécessité explicative (ANE)**

1. Les concepts sont les représentations mentales dont les propriétés expliquent le déroulement des processus cognitifs supérieurs.
2. Il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs.
3. Ces types de représentations mentales ont peu de propriétés en commun.
4. Donc, peu de propriétés sont communes aux représentations qui expliquent le déroulement des processus cognitifs supérieurs (de 2 et 3).
5. Donc, les concepts ne forment pas une espèce naturelle (de 1 et 4).

La prémisse 1 résulte du principe d'efficacité causale (chapitre 1, section 2.3.). La plupart des psychologues et des philosophes l'acceptent. Cependant, elle demeure controversée dans la mesure où le principe d'efficacité causale est vivement rejeté par Fodor. J'ai défendu le principe d'efficacité causale dans le premier chapitre. Je ne reviens pas sur ces arguments et je tiens ce principe pour vrai dans ce qui suit. Les chapitres suivants ont pour objet de défendre les prémisses 2 et 3 de l'ANE.

### 2.5. Conclusion

Dans cette section, j'ai appliqué la notion causale d'espèce naturelle à l'hypothèse EN. J'ai explicité mon rejet de cette hypothèse : les concepts ne forment pas une espèce naturelle, parce que les représentations qui figurent dans les processus cognitifs supérieurs humains n'ont pas de nombreuses propriétés en commun. En conséquence, cette classe de représentations mentales ne donne pas lieu à de nombreuses généralisations scientifiquement intéressantes. La raison est que ces représentations mentales se divisent en types qui sont très différents les uns des autres (l'hypothèse de l'hétérogénéité) : la plupart des catégories sont représentées par plusieurs concepts qui appartiennent à des types de représentations mentales qui ont peu à voir les uns avec les autres. L'argument de la nécessité explicative prétend établir cette affirmation. Le reste de ce travail a pour objet d'établir ses principales prémisses.

## 3. Propositions analogues

Dans la dernière section de ce chapitre, je compare mon hypothèse avec d'autres propositions. En effet, certains chercheurs ont exprimé des idées plus ou moins semblables. Cependant, la plupart de ces propositions diffèrent de manière substantielle des idées que je défends dans le présent travail. Cela peut expliquer pourquoi peu de psychologues ont conclu que les concepts ne formaient pas une espèce naturelle.

### 3.1. Diversité diachronique

On a souvent affirmé que les concepts varient *diachroniquement* (voir la discussion critique dans Keil et al. 1998). Certains psychologues ont ainsi soutenu que les concepts varient avec l'âge de leurs possesseurs (par exemple, Inhelder et Piaget 1969, Vygotsky 1986, voir la recension de Keil 1989 et le dogme 1 dans Keil et al. 1998). Selon cette proposition, des classes de représentations mentales fort différentes, voire entièrement différentes, figurent dans les processus cognitifs des adultes et des enfants (et, éventuellement, des enfants à différents âges).

D'autres psychologues ont soutenu que les concepts variaient fortement, voire entièrement, avec l'expérience (par exemple, Homa et al. 1981, voir le dogme 2 dans Keil

et al. 1998). Par exemple, certains ont affirmé que les individus qui savent peu de choses sur une catégorie utilisent des exemplaires, tandis que ceux qui ont davantage de connaissances utilisent des prototypes.<sup>17</sup> Ainsi, quelqu'un qui connaît peu de choses à propos des kangourous peut utiliser des souvenirs singuliers des rares individus qu'il a pu observer (par exemple, à la télévision). Au contraire, les Australiens qui sont familiers avec ces animaux peuvent utiliser des prototypes.

Dans les deux cas, les concepts sont supposés se diviser diachroniquement en types de représentations mentales qui ont peu en commun. Ma proposition est fort différente. Je soutiens *une diversité synchronique*. Je propose qu'au même moment, une catégorie est représentée par plusieurs concepts qui appartiennent à des types différents de représentations mentales. Ces types de représentations mentales ont peu en commun.

### 3.2. *Les théories composites des concepts*

On trouve assez souvent dans la littérature psychologique l'idée que les concepts sont faits *de parties hétérogènes, mais articulées*. Ces parties sont supposées encoder différents types de connaissances. Elles sont aussi supposées être activées par défaut dans des tâches cognitives différentes. Elles sont *articulées* en ce que les processus qui utilisent une partie peuvent être corrigés par les processus qui en utilisent une autre, mais non l'inverse (voir le chapitre 8 où je critique les théories composites des concepts).<sup>18</sup>

Par exemple, après la révolution des prototypes menée dans les années 70 par Posner, Smith et Rosch, de nombreux psychologues ont été attirés par *les théories duelles des concepts*. Ces théories soutiennent que les concepts sont constitués de deux parties, un cœur et une périphérie (voir Miller et Johnson-Laird 1976, Smith et Medin 1981, Osherson et Smith 1981, Armstrong et al. 1983). Le cœur du concept est une définition, tandis que sa périphérie est un prototype ou, plus généralement, une procédure de reconnaissance (R. Gelman (1990, 2001) présente une version actualisée de cette idée en faisant appel à la théorie des théories<sup>19</sup>). Ces parties sont supposées expliquer des phénomènes différents. Ainsi, selon Osherson et Smith (1981), le cœur définitionnel

---

<sup>17</sup> Sur ces notions, voir le chapitre suivant.

<sup>18</sup> Dans le chapitre 8, j'appelle cette propriété la *hiérarchisation fonctionnelle* des parties des concepts. Par exemple, les catégorisations qui font appel au prototype de grand-mère peuvent être corrigées sur la base de la définition de grand-mère, mais non l'inverse.

<sup>19</sup> Voir les chapitres 5 et 8.

figure dans des jugements réfléchis à propos de l'appartenance catégorielle des objets et est aussi impliqué dans la composition des concepts ; au contraire, le prototype est impliqué dans les jugements rapides d'appartenance catégorielle : il est utilisé en première approximation dans des catégorisations rapides et révisables.<sup>20</sup> Selon Miller et Johnson-Laird (1976), le cœur, mais pas la périphérie, est impliqué dans la compréhension du langage.

Les théories composites des concepts ressemblent à ma proposition.

- Elles postulent l'existence d'une certaine diversité conceptuelle : les diverses parties des concepts encodent différents types de connaissances.
- La plupart des catégories, ou au moins de nombreuses catégories, sont dénotées par un concept composé de plusieurs parties. Cette idée est analogue à la thèse que la plupart des catégories sont dénotées simultanément par plusieurs concepts qui appartiennent à différents types de représentations.

Cependant, je rejette les théories composites des concepts. Deux points essentiels distinguent ma proposition de ces théories :

- Les théories composites des concepts admettent habituellement que les parties des concepts sont employées dans différentes compétences cognitives (par exemple, Osherson et Smith 1981). La plupart des affirmations de ce type ont été falsifiées (voir la section 2.3. et le chapitre 4).
- Je rejette l'idée que les différents ensembles de connaissances à propos d'une classe soient essentiellement *articulés*. En conséquence, ces ensembles de connaissances doivent être conçus comme des représentations mentales différentes, et non comme des parties d'une unique représentation mentale (voir le chapitre 8).

Les théories composites des concepts feront l'objet d'une critique systématique dans le chapitre 8. Pour l'instant, notons seulement qu'elles sont foncièrement *instables*. En effet, les deux parties ne sont pas sur un pied d'égalité : l'une d'entre elles est subordonnée à l'autre. En conséquence, la partie subordonnée, par exemple le prototype dans la proposition d'Osherson et Smith (1981), ressemble davantage à un supplément

---

<sup>20</sup> Cette idée a été falsifiée (voir Murphy 2002, 24 sq.).

optionnel du concept qu'à une partie véritable du concept (les théories composites plus récentes, comme celle de Gelman, ont le même problème).<sup>21</sup>

### 3.3. *L'approche de Frank Keil*

Selon Frank Keil, il y aurait deux types de concepts (Keil 1989, 1991, Keil et al. 1998). Il distingue la classe *des concepts théoriques* de la classe *des concepts prototypiques*. Les concepts théoriques sont formés par des connaissances semblables à celles qui sont utilisées dans des explications scientifiques, tandis que les concepts prototypiques encodent des connaissances statistiques.<sup>22</sup>

Keil soutient que l'importance relative de ces deux types de concepts dans la cognition change durant le développement cognitif. Les jeunes enfants utilisent largement des concepts prototypiques. Ils possèdent des concepts théoriques, mais ceux-ci sont encore peu développés. Cependant, progressivement, ces concepts théoriques sont affinés et enrichis ; ils finissent par jouer un rôle central dans la cognition des enfants. Ce développement varie d'un domaine à l'autre (« domain-specific »). Keil écrit ainsi (1989, 18-19, voir aussi 23 et la conclusion 265 ainsi que Keil 1991) :

« I will argue that most concepts are intrinsically made up of both causal explanatory beliefs and sets of atheoretical relations on which the causal explanatory elements are only partially overlaid. (...) Although a child might have both representational types available to her, there may be radical shifts in the degree to which she relies on one or the other. For example, early on in a domain a child may have only the most skeletal framework of causal beliefs that guide the noticing of correlations, frequencies and so on. By contrast, an older child or adult might have the same conceptual space heavily infiltrated with causal beliefs (...). At all points in development the child has the competence for belief-laden representations and more associative ones; but for a given concept the balance between the two may change so dramatically that the dominant representational type does indeed shift. »

---

<sup>21</sup> En outre, la théorie d'Osherson et Smith hérite de toutes les difficultés des théories définitionnelles des concepts – voir le chapitre suivant.

<sup>22</sup> Voir le chapitre suivant.

Le livre de Keil (1989) s'efforce d'évaluer cette hypothèse.

Ma proposition ressemble à celle de Keil par certains aspects.

- Keil admet que la plupart des catégories, ou au moins un grand nombre d'entre elles, sont dénotées par plusieurs concepts qui appartiennent à des types différents.
- Keil admet aussi que chaque concept est utilisé par plusieurs capacités cognitives. Nous pouvons utiliser un prototype aussi bien quand nous catégorisons que quand nous induisons.

Cependant, ma position diffère de celle de Keil.

- Je crois qu'il y a plus de deux types hétérogènes de concepts. J'admets que les concepts théoriques et les prototypes sont deux types importants de concepts. Mais je soutiens que ces deux types n'épuisent pas la variété de nos concepts (voir le chapitre suivant).
- Keil admet une dépendance hiérarchique entre ces deux types de concepts. Chez les adultes, les concepts prototypiques dépendent des concepts théoriques (par exemple, Keil 1989, 265, Keil et al. 1998). Sur ce point, sa position est semblable aux théories composites des concepts. Par exemple, un concept théorique de chien et un concept prototypique de chien sont apparemment supposés être articulés d'une manière ou d'une autre.<sup>23</sup> Je rejette cette idée (chapitre 8).

### *3.4. Les représentations mentales déclaratives et non déclaratives*

Les psychologues et les chercheurs en neurosciences cognitives font souvent une distinction entre deux types de connaissance. La terminologie n'est pas complètement fixée : certains opposent la connaissance déclarative à la connaissance non déclarative (Knowlton 1997) ; d'autres distinguent la connaissance implicite de la connaissance

---

<sup>23</sup> Ainsi Johnson et Keil écrivent-ils (2000, 356) : « Explanatory relations are the contingencies that organize a concept, explaining and selecting feature correlations. These deep relations remain constant (within limits) as surface features can be incorporated. As conditional explanatory relations, they reflect the causal or intentional design of entities they represent. This design constrains the concept's interaction with others by requiring that goal-serving conditions be met and that established contingencies be maintained. Although superficial concept structures such as feature typicality and correlation and representative exemplars may often be relied on for diagnostic help, explanatory relations may underlie structures and direct their productive use. »

explicite (Goschke 1997, Shanks 1997b) ; en intelligence artificielle, certains distinguent la connaissance déclarative de la connaissance procédurale (voir la recension de Knowlton 1997). En outre, il n'est pas certain que les psychologues conçoivent tous cette distinction de la même manière (Clark 1993, chap. 6, Goschke 1997, Knowlton 1997). Parfois, des chercheurs qui utilisent la même terminologie semblent avoir à l'esprit des distinctions différentes.

Comme de nombreux psychologues, Knowlton distingue la connaissance qui est consciemment accessible – la connaissance déclarative – de celle qui ne l'est pas – la connaissance non déclarative (Knowlton 1997, voir le chapitre 5, section 4.3.). La connaissance est consciemment accessible si l'individu qui possède cette connaissance sait qu'il a acquis cette connaissance, même si le contenu exact de cette connaissance n'est pas consciemment accessible.<sup>24</sup> Sur la base des performances de patients amnésiques dans une variété de tâches expérimentales, Knowlton soutient que les êtres humains stockent les connaissances à propos des classes sous forme de représentations déclaratives *et* de représentations non déclaratives.<sup>25</sup> En effet, les patients amnésiques conservent la capacité d'acquérir des connaissances, alors même qu'ils sont supposés avoir perdu leur capacité d'acquérir des connaissances déclaratives.

On peut de prime abord douter que ces représentations non déclaratives soient bien des représentations conceptuelles, autrement dit, qu'elles figurent dans nos processus cognitifs supérieurs. Certains exemples de représentations non déclaratives conservées par les patients amnésiques renforcent le doute : par exemple, on souligne souvent le fait que les patients amnésiques conservent la capacité d'acquérir des compétences motrices (voir la recension de Knowlton 1997). Or, de prime abord, les habilités motrices ne sont pas conceptuelles (mais voir Prinz 2002, chapitre 5 et le chapitre 3, section 3. de ce travail). Cependant, d'autres types de représentations non déclaratives sont bien *conceptuels*. En particulier, certains patients sont à même

---

<sup>24</sup> La notion d'accessibilité à la conscience est donc prise en un sens particulier.

<sup>25</sup> Ou, plus précisément, dans au moins deux types de représentations, dans la mesure où Knowlton souligne correctement qu'il peut y avoir plusieurs types de représentations non déclaratives.

d'acquérir des connaissances utilisées dans des tâches de catégorisation (Knowlton et Squire 1993, Knowlton 1999, Reber et al. 1998, Smith et Minda 2000).<sup>26</sup>

Cette proposition est semblable à la mienne :

- Il y a différents types de représentations.
- De nombreuses catégories peuvent être dénotées simultanément par plusieurs concepts.
- Plusieurs compétences cognitives peuvent utiliser un même concept.

Il y a cependant des différences importantes entre ma proposition et celle de Knowlton :

- Je n'insiste pas sur la distinction entre les représentations déclaratives et les représentations non déclaratives, car il est possible, et à mes yeux probable, que cette distinction soit orthogonale à la manière la plus pertinente de classer nos représentations mentales. Par exemple, il est possible, et même probable, que la classe des représentations non déclaratives soit très hétérogène. Nous pouvons avoir des prototypes déclaratifs, des concepts théoriques déclaratifs, des définitions déclaratives etc.
- Knowlton ne montre pas que les connaissances déclaratives et non déclaratives ont peu en commun.
- Knowlton ne s'intéresse pas aux conséquences théoriques de sa position. Elle n'en tire aucune conclusion à propos de la notion de concept.

### *3.5. Les aspects originaux de ma position*

Je résume ce que je tiens pour les aspects les plus originaux de ma position :

- Certains psychologues ont reconnu une certaine forme de diversité parmi les concepts. Cependant, ils ont souvent interprété cette diversité au moyen des théories composites des concepts. Je vais longuement argumenter contre ces théories dans le chapitre 8.

---

<sup>26</sup> Quelle est la nature de ces représentations conceptuelles non déclaratives ? Knowlton propose que les représentations non déclaratives qui sont utilisées dans ces tâches de catégorisation sont des prototypes (voir le chapitre 5, section 4.1., et Knowlton et Squire 1993, Squire et Knowlton 1995, Knowlton 1997, 1999).

- La diversité entre types de concepts ne montre pas que l'hypothèse EN est fautive (section 2.1.). Je soutiens que les différents types de concepts ont peu à voir les uns avec les autres.
- Je conclus que les concepts ne forment pas une espèce naturelle. Cela va à l'encontre de l'image traditionnelle de l'esprit humain, admise aussi bien par les philosophes que par les psychologues. Les psychologues mentionnés ci-dessus n'ont pas tiré cette conclusion.

#### 4. Conclusion

Dans ce chapitre, j'ai explicité les thèses défendues dans le présent travail. Les concepts ne forment pas une espèce naturelle (*rejet de l'hypothèse EN*), parce qu'ils n'ont pas de nombreuses propriétés en commun. En effet, les représentations mentales dans les processus cognitifs supérieurs humains se divisent en types de représentations qui ont peu en commun (*l'hypothèse de l'hétérogénéité*). La plupart des catégories sont dénotées par plusieurs concepts qui appartiennent à ces types de représentations. *L'argument de la nécessité explicative* (l'ANE) vise à défendre l'hypothèse de l'hétérogénéité : il est justifié de postuler l'existence de types de représentations mentales très différents les uns des autres afin d'expliquer le déroulement des processus cognitifs supérieurs.

Des psychologues ont avancé des idées qui ressemblent au rejet de l'hypothèse EN : en particulier, ils ont reconnu l'existence d'une hétérogénéité conceptuelle. Mais la nature de cette hétérogénéité n'a pas été clairement déterminée. Ainsi, ces psychologues et ces philosophes n'ont pas défendu l'hypothèse de l'hétérogénéité ni le rejet de l'hypothèse EN.

Dans les chapitres suivants, je défends les prémisses de l'argument de la nécessité explicative. Si cette défense est convaincante, il faudra conclure que la notion de concept ne dénote pas une classe utile pour formuler des généralisations psychologiques : les représentations mentales qui sont utilisées par nos compétences cognitives supérieures ne constituent pas une classe scientifiquement importante.

## Chapitre 3

### Trois types de concepts

---

Dans le deuxième chapitre, j'ai explicité le rejet de l'hypothèse EN. Pour défendre ce rejet, j'ai présenté l'argument de la nécessité explicative.

#### **L'argument de la nécessité explicative (ANE)**

1. Les concepts sont les représentations mentales dont les propriétés expliquent le déroulement des processus cognitifs supérieurs.
2. Il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs.
3. Ces types de représentations mentales ont peu de propriétés en commun.
4. Donc, peu de propriétés sont communes aux représentations qui expliquent le déroulement des processus cognitifs supérieurs (de 2 et 3).
5. Donc, les concepts ne forment pas une espèce naturelle (de 1 et 4).

Dans ce chapitre et dans les chapitres suivants, je défends les prémisses de cet argument. L'idée centrale de l'ANE est exprimée par la prémisse 2 : nous devons postuler l'existence de plusieurs types de représentations mentales afin d'expliquer les propriétés des compétences cognitives supérieures des êtres humains. *Dans ce chapitre, je commence à défendre cette prémisse : je décris les types de représentations mentales qui, selon moi, doivent être acceptés dans notre ontologie, si l'on veut expliquer les propriétés des compétences cognitives supérieures.*<sup>1</sup>

Quels sont donc, selon moi, les types de représentations qui sont utilisés par nos compétences cognitives supérieures ? Tout d'abord, il doit être clair que la nature et le nombre de ces types de représentations est une question *empirique*. Il est donc impossible

---

<sup>1</sup> La défense de cette prémisse se poursuit dans les chapitres suivants : dans les chapitres 5 à 7, je présente des données empiriques en faveur de l'idée que ces différents types de représentations sont requis à des fins explicatives.

de déterminer *a priori* le nombre et la nature de ces types de représentations. Ceci posé, ma thèse est la suivante. Les principaux paradigmes théoriques en psychologie des concepts ont postulé l'existence de trois types de représentations, des prototypes, des exemplaires et des théories. *Alors que ces entités théoriques sont habituellement traitées comme des manières rivales de caractériser la classe des concepts, je propose que ce sont en fait trois types différents de concepts.* Dans ce chapitre, je présente ces trois entités théoriques. J'examine aussi les entités théoriques qui ont été postulées par d'autres théories des concepts.

Dans la première section, je décris les trois principaux paradigmes des concepts dans la psychologie contemporaine des concepts, à savoir le paradigme postulant des prototypes, le paradigme postulant des exemplaires et le paradigme postulant des théories. Dans la seconde section, je décris d'autres entités théoriques afin de déterminer si et dans quelle mesure elles sont différentes des entités postulées par les trois paradigmes principaux. Dans la troisième section, je présente en détail les entités théoriques postulées par Larry Barsalou et par Jesse Prinz, respectivement les symboles perceptuels et les proxytypes.

## **1. Prototypes, exemplaires et théories**

Depuis le rejet de ce que l'on appelle d'habitude le paradigme classique des concepts (PC) dans les années 70 (Posner et Keele 1968, 1970, Rips et al. 1973, Rosch 1973, 1975, 1978, Rosch et Mervis 1975, Rosch et al. 1976, Hampton 1979), trois paradigmes dominent la psychologie contemporaine des concepts, le paradigme postulant des prototypes (PP), le paradigme postulant des exemplaires (PE) et le paradigme postulant des théories (PT).<sup>2</sup> On les traite habituellement comme des théories rivales à propos de la classe prétendument homogène des concepts (voir, par exemple, Medin 1989, Komatsu 1992, Margolis 1994, Hampton 1997a, Laurence et Margolis 1999, Murphy 2002). La

---

<sup>2</sup> Le PP et le PE sont parfois assimilés sous le nom « d'approche fondée sur la ressemblance » (« similarity-based approach »). Le PT est aussi appelé parfois « théorie des théories » (« theory theory »), « l'approche de la connaissance » (« knowledge approach »), « l'approche fondée sur l'explication » (« the explanation-based approach »). Le nom « theory theory » est aussi utilisé dans le débat à propos des mécanismes cognitifs qui sous-tendent notre capacité de comprendre autrui (« mindreading »).

vérité de l'une de ces théories est supposée impliquer la fausseté des autres. Par exemple, Rips écrit à propos de l'article fondamental de Murphy et Medin (1985/2004) :

« If this conclusion is correct [that categorization is not a matter of similarity], it has important implications, *since it undercuts a whole class of models in cognitive psychology.* »<sup>3</sup>

Je soutiens que c'est une erreur : ces paradigmes portent en fait sur trois types différents de concepts dont il faut postuler l'existence pour expliquer les propriétés des compétences cognitives supérieures. Je propose aussi que ces paradigmes ont correctement caractérisé les principaux aspects de ces différents types de concepts. Dans ce qui suit, je décris les entités postulées par ces trois paradigmes.

Le lecteur doit avoir présent à l'esprit que ces paradigmes *ne sont pas* des théories détaillées des concepts. Chaque *paradigme* offre une conception générale de la nature des concepts. Cette conception peut être explicitée en détail par des *théories* spécifiques qui admettent les thèses de base du paradigme. Par exemple, à proprement parler, il n'existe pas *une* théorie des prototypes. Mais il y a une famille de théories des prototypes qui ont en commun des hypothèses semblables à propos de la nature des concepts, mais qui les développent différemment. C'est aussi le cas pour les autres paradigmes.

Dans ce qui suit, je me contente de décrire les thèses centrales des paradigmes. Pour défendre le rejet de l'hypothèse EN, je n'ai en effet pas besoin d'admettre une manière particulière de développer ces paradigmes. Comme d'autres les ont correctement décrits, je ne les présente pas en détail (voir Smith et Medin 1981, Medin 1989, Komatsu 1992, Margolis 1994, Hampton 1997a, Prinz 2002, chap. 2-4, Murphy 2002). Le lecteur intéressé peut consulter les recensions précédemment mentionnées.

### *1.1. Le paradigme classique des concepts (PC)*

---

<sup>3</sup> Je souligne.

Le PC a été l'approche dominante en psychologie des concepts depuis le début de la psychologie expérimentale jusqu'aux années 70 du vingtième siècle (Murphy 2002, chapitre 1, Machery<sup>4</sup>). Il repose sur l'idée que les concepts sont des *définitions* :

Un concept d'une classe d'objets est une définition, c'est-à-dire un ensemble de connaissances à propos des propriétés qui sont tenues pour nécessaires et conjointement suffisantes pour être un membre de cette classe (par exemple, Hull 1920, Katz et Fodor 1963).

Par exemple, posséder un concept de chien, c'est avoir un ensemble de connaissances à propos des propriétés qui sont tenues pour nécessaires et conjointement suffisantes pour être un chien.

Pour éviter toute incompréhension, le PC ne soutient pas que posséder un concept, c'est connaître la définition *réelle* de cette classe, c'est-à-dire les conditions réellement nécessaires et suffisantes pour appartenir à cette classe (à supposer qu'il y ait de telles conditions, voir Wittgenstein 1963 et Hull 1965). Le PC soutient qu'un concept est un ensemble de croyances sur les propriétés qui sont *traitées* par le possesseur du concept comme définissant la classe dénotée (voir Margolis 1994).<sup>5</sup> Un exemple peut éclairer cette distinction. Supposons que le concept d'aigle possédé par Pierre soit la définition suivante : « un objet est un aigle si et seulement s'il ressemble aux gravures sur les quarts<sup>6</sup> » (McNamara et Sternberg 1983). Le concept possédé par Pierre est une définition : elle stipule des conditions nécessaires et suffisantes pour être un aigle. Pierre fait appel (par défaut – voir le chapitre 1, section 1.2.3.) à cette définition quand il s'engage dans une tâche cognitive à propos des aigles. Par exemple, quand il catégorise un objet comme un aigle, il cherche à déterminer s'il ressemble aux gravures sur les pièces de 25 cents. De toute évidence, le concept de Pierre encode une fausse croyance : un oiseau peut être un aigle sans ressembler au profil gravé sur les pièces de 25 cents. Cependant, le concept de Pierre est bien une définition.

---

<sup>4</sup> Machery, E., « Concept theories : Old ideas in disguise », manuscrit.

<sup>5</sup> Cette distinction est semblable à la distinction faite par Locke entre essences nominales et essences réelles.

<sup>6</sup> Pièce américaine de 25 cents.

Les autres thèses de ce paradigme sont liées à cette idée centrale (Laurence et Margolis 1999). Par exemple, le PC s'accorde bien avec l'idée selon laquelle apprendre un concept, c'est abstraire un ensemble de propriétés qui caractérise tous les membres d'une classe et eux seuls (voir, par exemple, Fisher 1916 et Hull 1920). Pareillement, appliquer un concept à un individu, c'est-à-dire le catégoriser, c'est vérifier que cet individu possède toutes les propriétés nécessaires et conjointement suffisantes.

Le PC a été presque entièrement rejeté par les psychologues (Smith et Medin 1981, Laurence et Margolis 1999, Murphy 2002). Je ne vais pas faire la recension des raisons de ce rejet (Margolis 1994 en offre une évaluation critique intéressante). Aucune d'elles, considérée indépendamment, n'établit que les concepts humains ne sont pas des définitions ; cependant, prises ensembles, elles suggèrent fortement que le PC est une approche erronée pour la plupart des concepts.

Ce paradigme est encore défendu par certains psychologues et chercheurs en sciences cognitives – au moins pour certains concepts ou types de concepts (Jackendoff 1992a, Pinker et Prince 1999).<sup>7</sup> Cependant, même si le PC décrivait bien les représentations de certaines catégories, il peut être négligé dans ce qui suit. Car les données empiriques accumulées depuis trente ans montrent que *la plupart des catégories* ne sont pas représentées par un concept classique.

### *1.2. Le paradigme postulant des prototypes (PP)*

Le PP a été développé dans les années 70 pour remplacer le PC (Smith et Medin 1981, Margolis 1994, Laurence et Margolis 1999, Murphy 2002). Le PC a été très rapidement abandonné au profit du PP, en grande partie en raison des travaux d'Eleanor Rosch sur les représentations des catégories (Rosch 1973, Rosch et Mervis 1975, Rosch et al. 1976).<sup>8</sup> D'autres psychologues ont joué un rôle important dans l'apparition et le développement de cette approche, en particulier Michael Posner, Lance Rips, James

---

<sup>7</sup> Ainsi Pinker et Prince (1999) soutiennent-ils que nos concepts de relations familiales, nos concepts légaux et nos concepts moraux peuvent être adéquatement caractérisés par le PC, bien que d'autres types de concepts soient mieux caractérisés par le PP.

<sup>8</sup> Les idées de Rosch sur les concepts ont évolué. Elle a soutenu avec raison que ses résultats expérimentaux étaient compatibles avec plusieurs théories des concepts (Rosch 1978).

Hampton et la collègue de Rosch, Carolyn Mervis (Posner et Keele 1968, 1970, Rips et al. 1973, Hampton 1979, 1981).

Le PP repose sur l'idée que les concepts sont des prototypes<sup>9</sup> :

Un concept d'une classe d'objets est un *prototype*, c'est-à-dire un ensemble de connaissances statistiques sur les propriétés qui sont tenues pour être possédées par les membres de cette classe, en particulier sur les propriétés qui sont tenues pour être possédées par la plupart des membres de cette classe (Rosch et Mervis 1975, Hampton 1979).

Selon le PP, un concept est un ensemble de croyances statistiques : le possesseur du concept croit que les membres de la classe dénotée ont la probabilité  $x$  d'avoir la propriété  $P_1$ , la probabilité  $y$  d'avoir la propriété  $P_2$  etc.<sup>10</sup> Ces propriétés ne sont pas jugées nécessaires. En outre, dans la plupart des modèles des prototypes (mais pas dans tous), un prototype est un ensemble de croyances sur les propriétés qui sont jugées être possédées par *la plupart* des membres de la classe dénotée (les propriétés *typiques*).<sup>11</sup> Par exemple, posséder un concept de chien, c'est avoir une représentation d'une classe d'objets comme possédant en général, mais pas nécessairement, telle forme, comme étant en général composés de telles parties, comme ayant en général une queue...

Les prototypes ne représentent pas les propriétés qui sont *en fait* possédées par la plupart des membres de la classe dénotée. Le PP soutient que les propriétés représentées sont traitées comme si la plupart des membres de la classe dénotée les possédaient. Cependant, les avocats du PP admettent souvent l'idée que les prototypes sont appris par observation : l'apprenant observe des membres de la catégorie et détermine quelles propriétés sont les plus fréquentes (par exemple, Clark 1983, Hampton 2001). En

---

<sup>9</sup> Le terme « prototype » est utilisé de manière ambiguë pour désigner le meilleur exemplaire d'une catégorie ou la représentation d'une catégorie. J'utilise strictement le terme pour désigner la représentation mentale d'une catégorie.

<sup>10</sup> Cette mesure statistique peut n'être que qualitative: le possesseur du concept peut croire que *la plupart des* membres de la classe dénotée possède la propriété  $P$ .

<sup>11</sup> Certaines manières de caractériser le PP affaiblissent cette idée (par exemple, Prinz 2002, chapitre 3). En effet, certains affirment que n'importe quelle propriété peut être représentée par un prototype tant qu'elle n'est pas nécessaire. Par exemple, un prototype pourrait représenter les propriétés qu'un membre *idéal* de la classe dénotée posséderait. Je souligne au contraire l'idée que les prototypes sont des ensembles de connaissances statistiques. Les deux idées sont assez différentes (voir ci-dessous).

conséquence, si les objets que l'apprenant rencontre durant son apprentissage sont représentatifs de la catégorie entière, les propriétés qui sont représentées comme étant typiques peuvent en effet l'être.<sup>12</sup>

Je soutiens que le PP décrit les propriétés principales de l'un des types de concepts qui sont utilisés par les compétences cognitives supérieures des humains. Les êtres humains ont des ensembles de croyances statistiques, utilisés par défaut dans leurs processus cognitifs supérieurs, sur les propriétés qui sont possédées par les membres des classes dénotées. Cette thèse sera défendue dans les chapitres suivants (chapitres 5-7).

La plupart de catégories (catégories d'artéfacts, catégories d'animaux, catégories d'événements...) peuvent en principe être dénotées par un prototype (voir le chapitre 2, section 2.3., b.). Les partisans des prototypes ont accumulé des données empiriques qui sont compatibles avec cette possibilité. Les psychologues s'intéressent habituellement aux représentations des objets physiques de taille moyenne, par exemple aux chiens, aux tables, aux chaises... Cependant, certains psychologues ont aussi prêté attention à d'autres types de concepts. Ils y ont trouvé les phénomènes qui sont habituellement expliqués au moyen des prototypes (effets de typicalité, énumération spontanée de propriétés non nécessaires, corrélation entre la typicalité et le nombre de propriétés non nécessaires). Cela est compatible avec l'idée qu'un grand nombre de catégories dans un grand nombre de domaines sont représentées par des prototypes. Par exemple, Hampton (1981) a montré que c'est le cas de nombreux concepts abstraits, par exemple CROYANCE ou SCIENCE.<sup>13</sup> Les concepts d'événements pourraient aussi des prototypes : il est en effet possible d'identifier les scripts aux prototypes (Schank et Abelson 1974, Coleman et Kay 1981, Cantor et al. 1982). Un script peut être vu comme une description des parties typiques d'un événement. Les concepts sociaux comme

---

<sup>12</sup> Les êtres humains et les animaux en général semblent être de bons statisticiens (« statistical learners ») (Safran et al. 1996, Hauser et al. 2001). De toute évidence, tous nos concepts ne sont pas appris de cette manière : nous avons de nombreux concepts d'objets que nous n'avons jamais rencontrés ou bien que nous avons rencontrés une seule fois. Je ne discute pas cette idée dans le présent travail.

<sup>13</sup> Comme nous allons le voir, le PE peut aussi expliquer les phénomènes de typicalité (voir ci-dessous). En outre, la typicalité est elle-même un phénomène complexe qui a plusieurs sources (Armstrong et al. 1983, Barsalou 1985, Nosofsky 1988, Murphy 2002). En conséquence, les données empiriques de Hampton ne montrent pas décisivement que les concepts abstraits sont des prototypes. Mais elles montrent que les concepts abstraits pourraient être des prototypes et qu'ils sont probablement du même type que les concepts d'objets physiques.

INTROVERTI peuvent aussi être des prototypes (Cantor et Mischel 1979) : ici encore, on trouve les phénomènes qui caractérisent les concepts de catégories d'objets physiques comme CHAISE. C'est aussi le cas des concepts d'émotion (Shaver et al. 1992).

Les prototypes peuvent aussi en principe être utilisés par de nombreuses compétences conceptuelles (voir le principe de promiscuité dans le chapitre 1 et la section 2.3. a. du chapitre 2). Comme je le montrerai dans le prochain chapitre, les psychologues ont fait appel aux prototypes pour expliquer des propriétés communes à la plupart des processus cognitifs supérieurs, en particulier les propriétés de la catégorisation, de l'induction et de la déduction.

L'idée de base du PP peut être développée de multiples manières (Hampton 1979, 1998, Smith et al. 1988, Barsalou 1990). Par exemple, une théorie précise des prototypes devrait expliciter les points suivants.

- Quelle est la nature des propriétés représentées par les prototypes ? Par exemple, on pourrait avancer une théorie *empiriste* des prototypes, selon laquelle les prototypes représenteraient seulement les propriétés perceptuelles de leurs référents (Prinz 2002). Il importe cependant de noter que le PP n'est pas nécessairement empiriste (Hampton 1998, 138-139). Autrement dit, un prototype peut représenter n'importe quel type de propriétés. Ce qui distingue le PP des autres paradigmes, c'est que les prototypes sont des ensembles de croyances *statistiques* – non la nature des propriétés attribuées aux membres des classes dénotées.<sup>14</sup>
- Quel type d'information statistique est encodé par un prototype ? Est-ce que les prototypes encodent seulement la fréquence des propriétés ou encodent-ils aussi d'autres types d'informations statistiques ? Par exemple, ils pourraient encoder la fréquence des *cooccurrences* des propriétés (Barsalou 1990, Hampton 2001).
- Comment cette information statistique est-elle encodée ? Les informations statistiques peuvent être encodées de plusieurs manières (par exemple, via une moyenne ou une distribution de fréquences, voir Barsalou 1990).

---

<sup>14</sup> De nombreux critiques des théories des prototypes supposent à tort que les prototypes sont condamnés à représenter seulement des propriétés perceptuelles (par exemple, Keil 1989). Cela n'est pas le cas. Hampton écrit ainsi avec raison (1998, 138): « A moment's consideration shows that a model that is limited to representing purely perceptual information with no deeper structural, functional or abstract attributes is simply a 'straw man' as a model for representing concepts. »

- Quel est le véhicule des prototypes ?<sup>15</sup> Par exemple, Kosslyn a suggéré que les prototypes sont des images mentales (Kosslyn 1978). Est-ce le cas ? Sinon, que sont-ils ? Ces questions (et d'autres) sont cruciales si l'on veut développer une théorie spécifique des prototypes.<sup>16</sup> Cependant, je n'ai pas besoin de soutenir une version particulière du PP. J'ai seulement besoin d'affirmer l'existence de représentations qui encodent des connaissances statistiques à propos de leurs référents.

### 1.3. Le paradigme postulant des exemplaires (PE)

Peu après l'introduction du PP, certains psychologues ont proposé une autre approche des concepts, le paradigme postulant des exemplaires (Brooks 1978, Medin et Schaffer 1978).<sup>17</sup> Le PE n'a pas remplacé le PP, mais les deux approches se sont engagées dans une course expérimentale dans les années 80 (voir, par exemple, les numéros de *The Journal of Experimental Psychology: Learning, Memory, and Cognition* dans les années 80). Les psychologues des concepts ont cherché à trouver des données empiriques qui pourraient être expliquées par leur paradigme préféré, mais non par l'autre.

Le PE repose sur l'idée suivante : nos performances cognitives supérieures font appel à des représentations d'individus, qui sont utilisées comme des représentants de la catégorie entière.<sup>18</sup>

Un concept d'une classe d'objets est un ensemble d'exemplaires, c'est-à-dire de représentations de membres particuliers de cette classe.<sup>19</sup>

<sup>15</sup> Sur la notion de véhicule, voir le chapitre 1, section 1.1.2.

<sup>16</sup> Malheureusement, elles ne sont pas toujours traitées par les partisans des prototypes.

<sup>17</sup> Le modèle avancé par Medin et Schaffer (1978), appelé « le modèle contextuel » (« the context model »), est bien connu. Il a été étendu dans un autre modèle célèbre, le modèle contextuel généralisé (« the generalized context model ») (Nosofsky 1986, 1988, 1992).

<sup>18</sup> Cette idée est fort proche de la théorie berkeleyenne des idées générales (Berkeley 1734, Introduction, XVI). Berkeley écrit ainsi: « that though the idea I have in view whilst I make the demonstration, be, for instance, that of an isosceles rectangular triangle, whose sides are of a determinate length, I may nevertheless be certain it extends to all other rectilinear triangles, of what sort or bigness soever. And that, because neither the right angle, nor the equality, nor determinate length of the sides, are at all concerned in the demonstration. »

<sup>19</sup> Je tiens les exemplaires pour des représentations d'individus, et non des représentations de rencontres spécifiques avec des individus. Selon cette approche, le PE soutient que les êtres humains encodent une représentation pour chaque membre (représenté) d'une catégorie (Ross et Makin 1999). Pierre a un exemplaire de Fido, un exemplaire de Médor etc. Certains psychologues soutiennent au contraire que les exemplaires représentent des rencontres spécifiques avec des membres de la catégorie. Selon cette approche, chaque fois que Pierre rencontre Fido, il stocke un nouvel exemplaire (par exemple, Nosofsky

Par exemple, avoir un concept de chien, c'est avoir un ensemble de représentations de chiens particuliers, Pluto, Médor, Fido etc. et c'est utiliser ces représentations pour raisonner à propos des chiens en général. La classe des chiens est caractérisée comme la classe des entités qui sont semblables à Fido, Pluto et Médor.

Je soutiens que le paradigme postulant des exemplaires caractérise les propriétés principales de l'un des types de concepts. Nous représentons bien les catégories comme étant semblables à des individus spécifiques. Cette thèse sera défendue dans les chapitres suivants (chapitres 5-7).

Les exemplaires peuvent en principe représenter de nombreuses catégories dans de nombreux domaines. De toute évidence, les catégories d'objets physiques et d'événements peuvent être représentées par des exemplaires. On pourrait aussi représenter des entités abstraites, par exemple la justice et l'amour, au moyen d'événements spécifiques.<sup>20</sup>

En outre, les psychologues ont fait appel à ces entités théoriques pour expliquer de nombreuses compétences cognitives. Ils expliquent fort bien les propriétés de la catégorisation. En outre, de nombreuses psychologues soulignent l'importance des représentations d'individus, d'événements... singuliers dans notre cognition : selon eux, nous utilisons des connaissances à propos d'individus, d'événements ... singuliers pour la planification de nos actions, pour résoudre des problèmes etc. Cette approche est connue sous le nom de raisonnement fondé sur des cas (« case-based reasoning », voir Hammond 1989).<sup>21</sup> La notion d'exemplaire s'accorde de toute évidence avec cette approche des compétences cognitives supérieures.

---

1988, Barsalou 1990, Murphy 2002, 58 sq.). Pierre peut donc avoir de très nombreux exemplaires de Fido, un pour chaque rencontre avec Fido. Certaines données empiriques supportent la seconde version du PE (Nosofsky 1988, Murphy 2002, 58 sq.). Mais la première approche peut être modifiée pour les expliquer.

<sup>20</sup> Par exemple, notre concept de justice peut être un ensemble de représentations singulières de contrats ou d'échanges.

<sup>21</sup> Par exemple, nous utilisons des exemplaires pour raisonner par contre-exemple (Holyak et Glass 1975).

De nombreuses théories spécifiques peuvent développer ce paradigme (Nosofsky 1986, 1988, Barsalou 1990...). En particulier, une théorie complète des exemplaires devrait examiner les points suivants.

- Comment les membres particuliers sont-ils représentés ? Les représentations singulières ne peuvent représenter toutes les propriétés de leur référent (Smith et Medin 1981, Murphy et Medin 1985). Les théories des exemplaires doivent donc expliciter en détail quelles propriétés sont représentées par des exemplaires.
- Quelle est la nature du véhicule des exemplaires ? Les exemplaires sont-ils des images ? Sinon, que sont-ils ?

Je ne soutiens aucune théorie particulière des exemplaires. Pour l'argument défendu dans le présent travail, seul importe vraiment l'existence d'exemplaires.

#### *1.4. Le paradigme postulant des théories (PT)*

Le PT a été développé indépendamment au milieu des années 80 par Gregory Murphy et Douglas Medin (1985/2004) d'un côté et par Susan Carey (1985) de l'autre. Le travail de Murphy et Medin appartient à la psychologie de la catégorisation et de l'apprentissage des concepts (« concept learning »), c'est-à-dire le champ disciplinaire où sont nés le PP et le PE. Au contraire, Carey est une psychologue développementale. Le PT est rapidement devenu un paradigme influent dans ces deux disciplines (voir Rips 1989, Gopnik et Meltzoff 1997, Medin et Atran 1999, Murphy 2002). Particulièrement, le PT est de nos jours admis par la plupart des psychologues développementaux cognitifs (Carey 1985, Carey et Spelke 1994, Carey et Johnson 2000, Keil 1989, 1994, Spelke 1990, 1994, Spelke et al. 1992, R. Gelman 1990, 2001, Gopnik et Wellman 1994, Gopnik et Meltzoff 1997, Bloom 2000...).

Le PT est souvent présenté de manière un peu vague. Les psychologues affirment tantôt que les concepts sont des théories, tantôt qu'ils font essentiellement partie de théories (voir Murphy et Medin 1985, Rips 1995, Laurence et Margolis 1999 sur cette ambiguïté). En outre, la notion de théorie est elle-même plutôt vague. Malgré ces imprécisions, le PT peut être expliqué comme suit. Il repose essentiellement sur deux idées (Machery 2002, Prinz 2002) :

- Un concept d'une classe d'objets exprime un ensemble de propositions théoriques à propos de cette classe. Une proposition est théorique dans la mesure où elle peut servir de prémisse dans des *explications* des propriétés possédées par les membres de cette classe (Murphy et Medin 1985, Keil 1989, 1991, Rips 1995, Keil et Wilson 1999). Gelman et Coley écrivent ainsi (1991, 147):

« Theoretical kinds result from our attempts to understand the environment, not simply to organize and catalog it. Such categories hang together to the extent that they can explain and predict phenomena, instead of simply describing them. »

A quoi ressemblent ces propositions théoriques ? Des généralisations nomologiques, causales, génériques et/ou fonctionnelles semblent pouvoir figurer dans des explications. En conséquence, un concept théorique est supposé représenter des généralisations nomologiques, causales, génériques, et/ou fonctionnelles qui expliquent les propriétés des membres de son extension. Par exemple, posséder un concept de chien, c'est posséder une espèce de théorie à propos des chiens, c'est-à-dire avoir un ensemble de connaissances nomologiques, causales, génériques et/ou fonctionnelles à propos des chiens, par exemple la connaissance que les chiens ont un pelage pour se protéger du froid etc. J'appelle ces ensembles de connaissances *des mini-théories* (Rips 1995, Prinz 2002).

- Certaines catégories forment un domaine.<sup>22</sup> Par exemple, les catégories d'animaux, entre autres, la catégorie des chiens, la catégorie des chats etc., forment ensemble un domaine, à savoir le domaine de notre biologie naïve (ou une partie du domaine de notre biologie naïve). Or, nous sommes supposés posséder certaines connaissances à propos de chaque domaine (Keil 1989, Carey 1985, Carey et Spelke 1994, Gelman 1990, 2001, Heit 1997, 2001 etc.). Notre biologie naïve est l'ensemble de nos connaissances à propos du domaine des animaux et des plantes en général. Notre psychologie naïve est notre connaissance à propos du domaine des états mentaux en général. Ces connaissances sont souvent appelées *des théories naïves*. Le contenu des

---

<sup>22</sup> Malgré son importance, la notion de domaine est assez mal définie en psychologie cognitive. Elle est appliquée aussi bien aux états mentaux des êtres humains qu'aux échecs, aussi bien à l'ensemble des animaux qu'au sexe des poulets...

concepts qui relèvent d'un même domaine est supposé être déterminé par la théorie naïve de ce domaine. Par exemple, notre connaissance à propos des animaux en général, autrement dit notre biologie naïve, détermine quel type de croyances constitue nos concepts d'animaux, par exemple nos concepts de chien, de chat etc. (voir le chapitre 8). Pour utiliser un exemple simpliste, nous prêtons attention aux comportements des animaux parce que notre biologie naïve dit que les comportements sont des propriétés très importantes pour les animaux. J'appelle ces connaissances générales *des théories d'arrière-plan*.

La première idée est donc qu'un concept théorique d'une classe exprime des propositions théoriques à propos de cette classe. La seconde idée est donc que la nature des propositions qui sont exprimées par un concept théorique est déterminée par nos connaissances théoriques générales à propos du domaine qui inclut la classe à laquelle le concept réfère. Les psychologues développementaux sont souvent davantage intéressés par cette seconde idée que par la première (voir Keil 1989, Gopnik et Meltzoff 1997, Carey et Johnson 2000).

Bien que les partisans du PT n'aient pas distingué nettement ces deux idées, elles sont néanmoins différentes et logiquement indépendantes. On pourrait en effet admettre l'une des deux sans soutenir l'autre. Par exemple, on pourrait combiner la notion de prototype et les théories d'arrière-plan. Nos théories naïves pourraient spécifier à quelles propriétés l'apprenant statisticien doit prêter attention (par exemple, notre biologie naïve pourrait contenir l'instruction « prête attention aux comportements des animaux »). En conséquence, les prototypes encoderaient la fréquence de ces propriétés. Selon cette approche, il n'y aurait pas de mini-théorie (cette idée est appelée par Heit « l'explication de notre connaissance d'arrière-plan en termes de pondération » - « the weighting account of background knowledge », voir Heit 1997, 2001).

Pour les partisans du PT, les mini-théories et les théories d'arrière-plan sont liées de la manière suivante. Les mini-théories expriment des propositions nomologiques, génériques etc. à propos des catégories qu'elles dénotent. Cependant, différents types de propositions sont pertinents pour différents types de catégories : un concept théorique de matière, par exemple SABLE, un concept théorique d'animal, par exemple CHIEN, et un

concept théorique d'artéfact, par exemple CHAISE, n'expriment pas les mêmes types de propositions théoriques. Comment l'esprit peut-il donc savoir quel type de propositions est pertinent pour quel type de catégories ? Les psychologues admettent que nos théories naïves déterminent quels types de propositions sont exprimés par nos mini-théories : nos théories naïves à propos des matières, des animaux et des artéfacts déterminent quels types de généralisations sont respectivement exprimés par SABLE, CHIEN et CHAISE.

Par la suite, j'entends la notion de concept théorique (ou de théorie) comme suit. Les concepts théoriques expriment des propositions théoriques : autrement dit, ce sont des mini-théories. En outre, leur contenu est influencé par une théorie naïve d'arrière-plan. Je propose en effet que ces deux idées décrivent un type important de concepts.

Remarquons qu'en principe, on peut avoir des concepts théoriques pour une grande variété de catégories. L'esprit peut en effet encoder des propositions théoriques à propos de la plupart des catégories. Cela est le cas pour les catégories d'objets physiques de taille moyenne, pour les catégories d'événements... En outre, comme nous allons le voir, les psychologues ont fait appel aux concepts théoriques pour expliquer de nombreuses compétences cognitives.

Le PT peut être développé de plusieurs manières. En particulier, on peut formuler la notion de théorie de manière plus ou moins stricte. Gopnik et Meltzoff (1997, voir aussi Gopnik et Wellman 1994) ont développé une notion stricte de théorie : les théories sont semblables aux théories scientifiques. Ils soutiennent aussi que cette notion caractérise entre autres nos théories d'arrière-plan. D'autres rejettent cette approche (Keil 1994, Atran 1994) : la notion de théorie doit être prise dans un sens suffisamment large et l'analogie avec les théories scientifiques ne doit pas être poussée trop loin. Cependant, il est remarquable que peu de psychologues ont vraiment explicité ce qu'ils entendent par théorie (mais voir l'effort originel de Murphy et Medin 1985 et Gopnik et Wellman 1994).

### *1.5. Conclusion*

Je soutiens que les principaux paradigmes contemporains en psychologie des concepts ont identifié trois types majeurs de représentations qui figurent dans nos processus conceptuels. Dans cette section, j'ai présenté ces trois types, à savoir les prototypes, les exemplaires et les théories. Dans les chapitres qui suivent, je présenterai les nombreuses données empiriques qui suggèrent que ces trois types de représentations existent bel et bien et qu'ils sont utilisés dans nos compétences cognitives : nous avons un prototype de chien, un ensemble d'exemplaires de chiens et une théorie à propos des chiens.

## **2. Schémas et idéaux**

Dans cette section, je présente d'autres entités théoriques. Dans quelle mesure ces entités théoriques diffèrent-elles des entités examinées précédemment ?

### *2.1. Variations sur les trois paradigmes*

Certaines théories se contentent en fait d'explicitier d'une manière originale les thèses de base de l'un des trois grands paradigmes. C'est le cas de l'approche appelée essentialisme psychologique (« psychological essentialism ») (Medin et Ortony 1989, Gelman et Hirschfeld 1999, voir le chapitre 7). Selon cette approche, certains concepts (particulièrement, les concepts d'animaux et de plantes) ou, peut-être, tous les concepts contiennent la croyance que les catégories dénotées possèdent une essence. L'essentialisme psychologique est en fait une version du PT. Par exemple, un concept essentialiste de chien exprime la proposition que les chiens ont une essence en commun, que la nature de cette essence soit ou non spécifiée. S'il faut croire les psychologues qui étudient notre biologie naïve (Atran 1990, Medin et Atran 1999), cette proposition est liée à d'autres propositions, par exemple aux propositions que cette essence est transmise par reproduction, qu'elle n'est pas affectée par le milieu dans lequel un animal vit et grandit, qu'elle se trouve à l'intérieur des animaux... (voir Gelman et Wellman 1991). Ce sont là des propositions théoriques.

### *2.2. Théories mixtes*

D'autres théories mélangent les thèses fondamentales des trois paradigmes présentés précédemment. On trouve deux types de théories mixtes, les théories composites (voir Miller et Johnson-Laird 1976, Osherson et Smith 1981, et les chapitres 2 et 8 du présent travail etc.) et les théories des schémas (Rumerlhart 1980, Smith et al. 1988<sup>23</sup>).

J'ai déjà présenté les théories composites des concepts. Elles divisent les concepts en différentes parties (parfois appelées le cœur et la périphérie) qui correspondent aux entités postulées par les trois grands paradigmes. Par exemple, Osherson et Smith (1981) proposent que les concepts sont constitués d'une définition (le cœur) et d'un prototype (la périphérie).

Les théories des schémas sont un peu différentes. Comme les théories composites, elles soutiennent que les concepts encodent différents types de connaissances. Mais, contrairement aux théories composites, ces connaissances sont conjointement utilisées par nos compétences cognitives. Considérons un exemple. Le modèle que Smith et al. (1988) propose peut être vu comme une théorie des schémas. Selon leur modèle, un concept POMME a la forme suivante (cet exemple est bien sûr fictif) :

<b>Attributs</b>		<b>Propriétés</b>	
<i>Poids</i>	<i>Attribut</i>	<i>Poids</i>	<i>Propriété</i>
1	Couleur	25	Rouge
		5	Vert
		1	Marron
0,50	Forme	15	Rond
		0	Carré
		5	Cylindrique
0,25	Utilisation	20	Cru
		10	Tartes
		5	Compote
0,25	Texture	25	Lisse
		5	Inégale

<sup>23</sup> Smith et collègues (1988) semblent tenir leur modèle pour une version du PP (voir les commentaires dans Smith et Osherson 1989, 63). Cependant, on peut aussi y voir une théorie des schémas.

	1	Avec des bosses
<b>POMME</b>		

**Schéma 3.1.**

Ce concept encode de nombreuses connaissances. Tout d'abord, il caractérise explicitement *les types de propriétés* qu'une pomme possède : par exemple, une pomme a une couleur, une forme etc. Smith et collègues appellent ces types de propriétés des *attributs*. Deuxièmement, pour chaque attribut, le concept représente les *propriétés* qu'une pomme possède : par exemple, une pomme peut être rouge, verte ou marron. Le concept représente aussi la fréquence relative de ces propriétés : la différence entre les poids des propriétés encode leur fréquence relative. Par exemple, les pommes vertes sont cinq fois plus fréquentes que les pommes marron. Finalement, les attributs sont aussi pondérés. Ces poids représentent l'importance de chaque propriété quand l'on catégorise un objet.

Dans ce modèle, comme dans le PP, les concepts encodent des connaissances statistiques à propos de leurs référents. Mais ils encodent davantage. En effet, ils encodent des connaissances sur l'importance des différents attributs pour catégoriser un objet. Le concept POMME représenté ci-dessus dit ainsi que pour décider si quelque chose est une pomme, on devrait prêter plus d'attention à sa forme qu'à son utilisation. Selon une interprétation possible, cette connaissance n'est pas statistique : elle ne dit rien sur la fréquence de telle et telle propriété. Elle est au contraire théorique : elle peut reposer sur nos connaissances d'arrière-plan à propos des plantes et des fruits. Un fruit, par exemple une pomme, a naturellement une forme typique, mais son utilisation n'est pas spécifique, puisque les poires, les pêches etc. sont utilisées de manière semblable.

Les schémas encodent donc plusieurs types de connaissances, statistiques et non statistiques. Mais ces types de connaissances ne sont pas utilisés par différentes compétences cognitives, comme dans les théories composites des concepts. Les différents types de connaissances sont utilisés conjointement par chaque compétence cognitive. Par exemple, dans le modèle de Smith et al., les connaissances à propos des attributs et les connaissances à propos des propriétés sont conjointement utilisées dans les décisions de catégorisation.

Je n'endors pas les théories des schémas. En effet, nous verrons dans le chapitre 5 que les différents types de connaissances ne sont souvent pas utilisés conjointement dans les catégorisations.

### 2.3. Une entité théorique originale : les idéaux

Certains psychologues ont proposé que la représentation d'une catégorie est en fait la représentation d'un membre *idéal* de la catégorie. Cette position est souvent confondue avec les théories inspirées par le PP, bien que les deux approches soient en fait différentes. Si les concepts sont des idéaux, ils ne représentent pas les propriétés que *la plupart* des chiens possèdent et, plus généralement, ils ne représentent pas une information *statistique* à propos des chiens. Ils représentent les propriétés qu'un chien parfait posséderait.

Certaines données empiriques, par exemple les travaux de Barsalou sur les représentations des catégories *ad hoc*, viennent à l'appui de cette approche (Barsalou 1983). Les catégories *ad hoc* sont des catégories dont les représentations ne sont pas stockées de manière permanente en MLT. Elles sont au contraire produites sur demande. Un exemple célèbre de catégories *ad hoc* est la classe des objets à emporter de sa maison en feu. Ces catégories ne sont pas représentées par des prototypes, puisque les membres de ces classes ont peu de propriétés en commun. Au contraire, ces catégories semblent être organisées autour d'idéaux. Barsalou a suggéré que cela était peut-être aussi le cas d'autres catégories (Barsalou 1985).

Il se peut en effet que nous ayons en MLT des représentations d'entités idéales et que nous les utilisions dans des tâches cognitives supérieures. Je ne vais cependant pas discuter en détail cette possibilité. Si nous avons bien de telles représentations et si elles sont bel et bien différentes des autres types de représentations, l'ANE en est d'autant renforcé. Ce qui m'importe avant tout dans ce travail.

### 2.4. Conclusion

Chacun des trois principaux paradigmes des concepts a posé l'existence d'un type particulier de représentations, les prototypes, les exemplaires et les théories. Dans cette section, j'ai présenté d'autres types de représentations. Certains ne diffèrent pas vraiment

des entités théoriques postulées par les trois paradigmes principaux. D'autres sont en fait assez différents. C'est le cas des entités postulées par les théories composites, des schémas et des idéaux de Barsalou. J'ai rejeté les deux premières entités théoriques et j'ai admis que les idéaux peuvent exister. Mais, si les idéaux existaient bel et bien, l'ANE en sortirait renforcé et non affaibli.

### **3. Proxytypes et symboles perceptuels**

#### *3.1. Les symboles perceptuels et les proxytypes*

Les *symboles perceptuels* de Barsalou et les *proxytypes* de Prinz offrent apparemment une alternative aux entités postulées par les trois principaux paradigmes théoriques (Goldstone et Barsalou 1998, Barsalou 1999, Barsalou et al. 2003b, Prinz 2002, voir aussi Damasio 1994 sur les zones de convergence). Les idées de Barsalou et Prinz se ressemblent. Elles peuvent être résumées comme suit :

- Barsalou et Prinz défendent une approche empiriste des concepts :
  - Les concepts sont *des états perceptuels* : extraire un concept de la MLT, c'est revivre (« reenact ») – souvent inconsciemment – les états perceptuels dans lesquels le sujet se trouvait quand il percevait un membre de la classe dénotée par le concept. Ainsi, un individu a un concept quand il est en mesure d'utiliser dans ses processus cognitifs supérieurs les traces mnésiques de ses états perceptuels.
  - Toute notre connaissance, y compris notre connaissance abstraite, est stockée dans des états perceptuels spécifiques à chaque modalité.
- Les concepts sont sensibles au contexte : seul un sous-ensemble des connaissances stockées en MLT, déterminé en fonction du contexte d'extraction, est extrait de la MLT quand nous nous engageons dans une tâche cognitive à propos de la classe représentée.

Ainsi Prinz écrit-il (2002, 149-150) :

« (...) concepts are proxytypes, where proxytypes are perceptually derived representations that can be recruited by working memory to represent a category.

A proxytype can be a detailed multimodal representation, a single visual object model, or even a mental representation of a word (e.g., an auditory image of the word « dog »). Every long-term memory network of perceptual representations contains many overlapping proxytypes. Just about any concise subset from such a network can do the job. Context determines what proxytype is used in working memory on any given occasion. If one is looking for dogs in the arctic tundra, one can call a representation of a typical sled dog. If one is looking for a guard dog, one can use a representation of a more ferocious breed. Sometimes these proxytypes already exist in long-term memory and sometimes they must be constructed. If one reads a news report about a dog that is 5 feet tall, one probably constructs a new representation on the fly (see Barsalou 1987). This flexibility makes proxytypes ideal for tracking things. »<sup>24</sup>

Ces idées sont extrêmement intéressantes, et de nombreuses données viennent à leur appui (voir les articles de Goldstone et Barsalou 1998, Barsalou 1999, Barsalou et al. 2003a, b, Prinz 2002, voir aussi Damasio 1994). L'hypothèse selon laquelle nous utiliserions des connaissances perceptuelles dans des tâches cognitives supérieures explique des propriétés spécifiques de nos processus cognitifs (Goldstone et Barsalou 1998, Barsalou et al. 2003a, b, Simmons et al. 2003). Par exemple, quand nous devons vérifier si un objet possède une propriété (par exemple, est-ce que les crabes ont des nageoires ?), nous visualisons cet objet.

### *3.2. Un défi pour le rejet de l'hypothèse EN*

Je ne vais pas contester ces données empiriques, car cela m'écarterait de ma ligne argumentative. Cependant, dans le présent contexte, ces idées soulèvent la question suivante : si j'admets ces données empiriques, est-ce que l'hypothèse EN en sort renforcée ? En fait, la proposition empiriste de Barsalou et Prinz va à l'encontre de ma

---

<sup>24</sup> Barsalou et collègues écrivent aussi (2003b, 85): «When a car is perceived visually, a set of neural feature detectors becomes active in the visual system (...). Conjunctive neurons in a nearby association area then conjoin the active features and store them in memory. Later, in the absence of visual input, these conjunctive neurons partially reactivate the original set of feature detectors to represent the car visually. Such reenactments or simulations are never complete and may be biased. Nevertheless they provide the cognitive-level representations that support memory, language, and thought. »

position. Car, selon Barsalou et Prinz, ces données montrent que contrairement à ce qu'ont postulé toutes les théories des concepts, les concepts doivent être *en général* identifiés à des proxytypes ou à des symboles perceptuels. Il y a donc bien de nombreuses propriétés communes à tous les concepts.

Répondons à cette thèse. Tout d'abord, quelle est exactement la différence entre les entités théoriques postulées par Barsalou et Prinz et les entités théoriques postulées par le PP et le PE ? Bien que Prinz et Barsalou ne soient pas parfaitement clairs sur ce point, leurs positions combinent trois éléments : une théorie mixte à propos du contenu des concepts (voir la section 2 de ce chapitre sur la nature des théories mixtes des concepts), une approche contextualiste de l'extraction des connaissances stockées en mémoire et une hypothèse néo-empiriste à propos du véhicule des concepts. En effet, pour ce qui est du contenu des concepts, leurs théories peuvent en effet être vues comme une version mixte du PP et du PE.<sup>25</sup> Les symboles perceptuels et les proxytypes sont apparemment supposés encoder à la fois des connaissances à propos d'individus singuliers, comme le font les exemplaires, et des connaissances statistiques, comme le font les prototypes. Barsalou et Prinz ajoutent aux théories mixtes l'idée originale que ces connaissances sont stockées dans un format représentationnel particulier, à savoir dans des états perceptuels spécifiques à chaque modalité. C'est une hypothèse à propos des *véhicules* de nos représentations mentales. C'est l'élément crucial et novateur de leur approche.

Or, je soutiens que les théories mixtes doivent être rejetées (voir la section 2.2. et le chapitre 8). Ainsi, je rejette l'existence de proxytypes et de symboles perceptuels, parce que je rejette les théories mixtes en général.<sup>26</sup>

Barsalou et Prinz pourraient peut-être avancer l'objection suivante. Il pourrait renoncer à leur théorie mixte et reconnaître que chaque catégorie est dénotée par plusieurs concepts. Mais ils pourraient soutenir que leurs données expérimentales montrent que ces ensembles de connaissances ne sont pas si différents que cela. Car ils

---

<sup>25</sup> Malgré la rhétorique de Barsalou, de nombreux aspects de sa théorie sont inspirés par le PP et le PE. Nombre des distinctions qu'il suggère entre sa théorie et les théories antérieures sont artificielles. Prinz est plus lucide (Prinz 2002).

<sup>26</sup> Cependant, comme nous allons le voir, j'accepte l'idée centrale des théories de Prinz et de Barsalou. Nous encodons bien des connaissances perceptuelles et nous nous servons de ces connaissances dans des tâches conceptuelles.

ont le même format représentationnel. Bref, ils pourraient soutenir que les exemplaires et les prototypes encodent les uns et les autres des connaissances dans un format perceptuel. En conséquence, l'on pourrait formuler des généralisations inductives importantes à propos de la classe des concepts en général. *L'hypothèse EN ne serait donc pas fausse.* En conséquence, en avançant une hypothèse empiriste à propos de la nature des véhicules des concepts, les théories de Prinz et de Barsalou viennent à l'appui de l'hypothèse EN.

Cependant, les données empiriques avancées par Prinz et Barsalou montrent seulement que nous avons *un* type de représentations qui satisfait l'élément central des théories de Prinz et de Barsalou, à savoir l'encodage perceptuel des connaissances conceptuelles. Mais cela ne revient en aucun cas à montrer que les concepts satisfont *en général* cette hypothèse. Tirer cette conclusion, c'est *admettre l'hypothèse EN* – sans la discuter (mais voir Goldstone et Barsalou 1998 qui sont plus prudents). Les données empiriques avancées par Prinz et Barsalou sont aussi compatibles avec la thèse suivante : les humains encodent de manière perceptuelle *certaines* informations et les utilisent dans les tâches conceptuelles. Il se peut donc qu'il y ait un type de concepts perceptuels en sus des autres types de représentations. Bien sûr, ce type de concepts pourrait être assez différent des autres types de concepts : il pourrait avoir un type particulier de format représentationnel, probablement un type spécifique de contenu et des propriétés fonctionnelles spécifiques (impliquant les systèmes perceptuels). *Cela renforcerait bien sûr le rejet de l'hypothèse EN.*

Barsalou et Prinz ne seraient pas satisfaits par cette réponse. Car ils croient qu'un seul format représentationnel, de nature perceptuelle, est suffisant pour encoder toutes les connaissances stockées en mémoire à long terme (Barsalou 1999, Prinz 2002). En conséquence, ils avancent un argument fondé sur le principe de parcimonie (Prinz 2002, Barsalou 1999, Barsalou et al. 2003b) : nous savons que nous encodons certaines connaissances dans des représentations perceptuelles ; toutes les connaissances peuvent être encodées dans ce format ; nous n'avons donc aucune raison de supposer l'existence d'autres formats représentationnels.

Pour défendre la seconde prémisse de ce raisonnement, Barsalou et Prinz s'efforcent de montrer comment un format perceptuel pourrait encoder différents types de connaissances, particulièrement des connaissances abstraites. Cependant, leurs

suggestions sont peu convaincantes. En outre, ni l'un ni l'autre n'expliquent pourquoi des connaissances de type théorique (sur la notion de théorie, voir la section 1.3. ci-dessus) affectent nos performances cognitives. Comme les connaissances abstraites, ce type de connaissances semble peu compatible avec un format perceptuel (même si Barsalou et Prinz nieraient probablement ce point). L'hypothèse nulle est donc que nous encodons de manière non perceptuelle des connaissances que nous utilisons dans nos performances cognitives supérieures. Je conclus que l'argument fondé sur le principe de parcimonie n'est pas convaincant. En conséquence, les données empiriques avancées par Barsalou et par Prinz ne suggèrent pas que les concepts forment une espèce naturelle.

### 3.3. Les concepts incarnés (« embodied concepts »)

Barsalou et Prinz avancent souvent deux idées en même temps :

- *Empirisme* : (i) le format représentationnel est perceptuel et (ii) extraire un concept de la MLT, c'est revivre un état perceptuel.
- *Incarnation* (« embodiment ») : nos concepts n'encodent pas seulement des connaissances à propos des propriétés de leurs référents, mais aussi des connaissances à propos de la manière dont nous sommes *situés* par rapport à eux. Ils encodent en particulier des connaissances à propos de nos interactions physiques avec leurs référents et à propos des réactions émotionnelles suscitées par leurs référents (Prinz 2002, chap. 5, Barsalou et al. 2003a).

L'association de ces deux idées n'est nullement arbitraire. Car la perception peut être caractérisée comme un type d'interaction avec les objets qui affectent nos sens (Noe et O'Regan 2001). Néanmoins, elles ne sont pas identiques : selon la seconde idée seulement, nos concepts encodent *des connaissances motrices*, c'est-à-dire des connaissances à propos de la manière dont nous interagissons physiquement avec les objets, et *des connaissances émotionnelles*, c'est-à-dire des connaissances à propos des nos réactions émotionnelles.<sup>27</sup>

---

<sup>27</sup> Voir Barsalou et al. 2003a.

L'idée de concepts incarnés est originale. Car les connaissances qu'ils encodent diffèrent du type de connaissances que les prototypes, les théories et les exemplaires sont supposés encoder. Les prototypes, les théories et les exemplaires sont en effet supposés encoder des connaissances à propos des propriétés *de leurs référents*.<sup>28</sup> Au contraire, les concepts incarnés encodent des connaissances à propos des agents, la manière dont ils interagissent avec les objets et la manière dont ceux-ci les affectent. Par exemple, Prinz fait la suggestion suivante (2002, 121, voir aussi Barsalou 2003a) :

« our knowledge of objects often includes knowledge about how we physically interact with them. Knowing what a hammer is involves knowing what to do with it, and knowing what to do with it involves knowing a sequence of motor movements. »

Ces spéculations sont maintenant supportées par un nombre croissant de données empiriques (Barsalou et al. 2003a, b). On a montré que les performances de catégorisation activent nos représentations motrices. Nos performances dans des tâches conceptuelles sont affectées par les mouvements que nous devons accomplir et par notre état émotionnel.

J'admets l'idée principale des spéculations de Barsalou et de Prinz. Il est plausible que nous stockions en MLT des connaissances motrices et émotionnelles et que nous les utilisions régulièrement dans nos processus cognitifs supérieurs.<sup>29</sup> Prinz et Barsalou proposent que ces informations constituent une *partie* de nos concepts. Comme je rejette les théories mixtes des concepts, je soutiens qu'elles constituent peut-être *un type particulier de concepts*, qui peut être assez différent des autres types de concepts.

### 3.4. Conclusion

---

<sup>28</sup> Néanmoins, ces conceptions pourraient être élargies. Par exemple, un prototype d'une classe x pourrait aussi encoder les propriétés fréquentes de nos interactions avec les membres de x.

<sup>29</sup> *Les informations incarnées doivent figurer par défaut dans nos processus conceptuels pour être véritablement conceptuelles – au moins dans certains contextes (voir chapitre 1).* En effet, les processus cognitifs supérieurs pourraient à l'occasion faire appel à des connaissances perceptuelles, motrices et émotionnelles sans que, *ipso facto*, ces connaissances soient conceptuelles.

Prinz et Barsalou soutiennent que les concepts doivent être considérés comme des proxytypes ou comme des symboles perceptuels. Leurs théories ne sont pas entièrement novatrices : ces entités théoriques sont, tout bien considéré, semblables aux schémas ou aux entités postulées par les théories composites des concepts. Cependant, leur thèse principale est plus originale : le format représentationnel de nos concepts est perceptuel. Comme nous l'avons vu, cela implique que les concepts forment une espèce naturelle. Cependant, même si j'endossais toutes les données empiriques avancées par Barsalou et par Prinz, je devrais conclure que nous encodons *certaines* connaissances de manière perceptuelle. Cela est parfaitement compatible avec le rejet de l'hypothèse EN : nous pouvons avoir un type de concepts perceptuels en sus des autres types de concepts. Barsalou et Prinz soutiennent aussi que nous encodons et utilisons certaines connaissances motrices et émotionnelles. Cela est peut-être vrai. Mais, puisque je rejette les théories mixtes des concepts, je propose qu'il conviendrait de postuler un autre type de concepts. Si ce type de concepts était très différent des autres types, cela renforcerait en fait le rejet de l'hypothèse EN.

#### **4. Conclusion**

Dans ce chapitre, j'ai commencé à défendre la prémisse 2 de l'ANE. La prémisse 2 soutient qu'il est nécessaire de postuler plusieurs types de représentations afin d'expliquer nos compétences cognitives supérieures. Dans ce chapitre, j'ai décrit les types de représentations qu'il faut postuler. Les trois grands paradigmes des concepts, la conception des prototypes, la conception des exemplaires et la conception des théories, ont postulé trois types d'entités théoriques, les prototypes, les exemplaires et les théories :

- les prototypes encodent des connaissances statistiques à propos de ce qu'ils représentent ;
- les exemplaires encodent des connaissances à propos d'individus ;
- les théories encodent des connaissances causales, fonctionnelles et/ou nomologiques.

Ces trois entités sont habituellement traitées comme des caractérisations rivales de la classe des concepts. Dans les chapitres 5 à 8, je cherche au contraire à montrer que ces trois types de représentations existent conjointement. Je soutiens donc que le PP, le PE et le PT décrivent les propriétés principales de trois types de représentations qui figurent dans nos processus cognitifs supérieurs.

J'ai aussi examiné d'autres théories. Certaines sont simplement des variantes de ces trois paradigmes. Cependant, les thèses empiristes de Barsalou et de Prinz sont assez différentes. Elles sont problématiques pour ma position. Car la thèse que le véhicule des concepts est de nature perceptuelle supporte l'hypothèse EN : il y aurait alors un unique format représentationnel commun aux différents types de concepts. Cependant, on peut à juste titre douter de l'idée que toutes les connaissances stockées en MLT sont encodées dans un format perceptuel. Au contraire, les données empiriques avancées par Barsalou et par Prinz suggèrent plutôt que nous avons un type particulier, peut-être *sui generis*, de concepts perceptuels. Cela est parfaitement compatible avec le rejet de l'hypothèse EN. En outre, Barsalou et Prinz proposent que nous utilisons de manière systématique des connaissances motrices et émotionnelles dans nos performances cognitives supérieures. J'endosse cette idée : nous pouvons en effet avoir des concepts incarnés. Cependant, de nouveau, si ce type de concepts est *sui generis*, cela renforce l'hypothèse de l'hétérogénéité.

Il est donc possible que nous ayons un type de concepts perceptuels et un type de concepts incarnés. Si ces concepts ont des propriétés spécifiques, cela renforce l'hypothèse EN. Cependant, par souci de simplicité, je vais prêter exclusivement attention aux trois entités principales, les prototypes, les exemplaires et les théories. Je compte montrer que nous possédons ces trois types de concepts et qu'ils ont peu en commun. C'est suffisant pour rejeter l'hypothèse EN.

## Chapitre 4

### L'hétérogénéité conceptuelle

---

Dans le chapitre 3, j'ai décrit les principaux paradigmes des concepts. Les entités qu'ils postulent sont de toute évidence différentes les unes des autres (voir, par exemple, la citation de Murphy dans le chapitre 1, section 2.4.). Cependant, l'argument de la nécessité explicative (ANE) requiert davantage. La prémisse trois soutient donc que les types de représentations qui sont nécessaires pour expliquer nos compétences cognitives supérieures ont *peu en commun* les uns avec les autres.<sup>1</sup> *Dans ce chapitre, je défends cette prémisse.*

Dans la première section, j'explique ce que l'on veut dire quand l'on soutient que deux types de représentations ont peu de propriétés en commun. Dans la deuxième section, je montre que les prototypes, les exemplaires et les théories sont supposés avoir des propriétés sémantiques différentes ainsi que des propriétés fonctionnelles différentes. Je conclus que ces trois entités théoriques ont peu de propriétés en commun. Dans la troisième section, je présente la théorie pluraliste des processus cognitifs. Dans la dernière section, je discute un argument développé par le psychologue Larry Barsalou, qui revient à nier la prémisse 3 de l'ANE (Barsalou 1990).

#### 1. Hétérogénéité conceptuelle

Pour rejeter l'hypothèse EN, je dois montrer que les concepts se divisent en types qui ont peu en commun. Dans cette section, j'explique ce que l'on veut dire quand l'on soutient que deux types de représentations mentales ont peu de propriétés en commun.

##### 1.1. *Quand deux types de représentations mentales ont-ils peu en commun ?*

Trivialement, les prototypes, les exemplaires et les théories ont peu de propriétés en commun si les propriétés qui sont généralement vraies des prototypes, des exemplaires et des théories sont différentes. Toutefois, toutes les propriétés ne sont pas pertinentes pour

---

<sup>1</sup> Pascal Ludwig et Luc Faucher ont insisté à juste titre sur ce point.

mesurer le degré de ressemblance de ces types de représentations. Par exemple, la propriété *ne pas être un chien* n'est possédée ni par les prototypes ni par les exemplaires. Mais de toute évidence, ce fait est dépourvu de pertinence pour évaluer la prémisse 3.

Quelles propriétés sont donc pertinentes ? Dire que les prototypes, les exemplaires et les théories ont peu en commun, c'est dire qu'ils ont peu en commun *pour des représentations*.<sup>2</sup> Le degré de ressemblance entre ces trois entités théoriques dépend donc des propriétés que nous tenons pour pertinentes pour caractériser un type de représentations mentales. Or, il est largement admis que trois types de propriétés sont importants pour caractériser des représentations mentales :

- *Leurs propriétés sémantiques* : quelles propositions sont exprimées par les concepts ?
- *Leurs propriétés fonctionnelles* : comment les concepts déterminent-ils le déroulement des processus cognitifs dans lesquels ils figurent ?
- *Les propriétés de leurs véhicules* : comment les concepts représentent-ils ce qu'ils représentent ?

Les deux premiers types de propriétés correspondent à la caractérisation fonctionnaliste des états mentaux (Putnam 1967/1975). Je me concentre sur eux. Car il n'y a pas encore de théorie développée à propos des véhicules des prototypes, des exemplaires et des théories (voir la discussion des théories empiristes de Barsalou et de Prinz dans le chapitre 3, section 3.).

### *1.2. La prémisse 3 est révisable*

L'évaluation de la ressemblance entre prototypes, exemplaires et théories est révisable. Car nous pourrions modifier notre théorie sur les propriétés que nous jugeons être importantes pour caractériser des types de représentations. On pourrait par exemple être amené à inclure les propriétés neuronales des représentations parmi ces propriétés. Supposons ainsi que les propriétés neuronales soient importantes pour expliquer de nombreux processus conceptuels : par exemple, la localisation neuronale des concepts pourrait expliquer les propriétés temporelles des processus cognitifs supérieurs (« time

---

<sup>2</sup> Par analogie, dire que les pêches et les poires ont peu en commun, c'est dire qu'elles ont peu en commun pour des fruits.

course »). Dans ces conditions, nous devrions modifier notre théorie sur les propriétés qui sont importantes pour caractériser un type de représentations et inclure la localisation neuronale des représentations parmi ces propriétés. Supposons aussi que les états neuronaux qui réalisent les prototypes, les exemplaires et les théories aient de nombreuses propriétés en commun : par exemple, ils pourraient tous être situés dans la même partie du cerveau, par exemple dans le cortex préfrontal. Nous devrions alors admettre que des propriétés pertinentes sont communes aux différents types de concepts et que les concepts forment une espèce naturelle. Le rejet de l'hypothèse EN dépend donc de l'état actuel de notre connaissance ; de ce fait, il est susceptible d'être révisé.

## **2. Les propriétés sémantiques et fonctionnelles des prototypes, des exemplaires et des théories**

### *2.1. Ces trois entités théoriques ont des propriétés sémantiques différentes*

On peut caractériser les différences sémantiques entre les représentations postulées par les trois principaux paradigmes des concepts comme suit.<sup>3</sup>

- Un *exemplaire* exprime des propositions qui attribuent des propriétés à un individu spécifique, c'est-à-dire des propositions dont la forme est [Fa]. Par exemple, un exemplaire de chien pourrait exprimer la proposition [Fido a une queue].
- Un *prototype* exprime des propositions statistiques. Habituellement, un prototype est supposé exprimer des propositions qui attribuent des propriétés à la plupart des membres de la classe dénotée, c'est-à-dire des propositions dont la forme est [pour la plupart des x, Fx]. Par exemple, un prototype de chien pourrait exprimer la proposition [la plupart des chiens aboient]. Un prototype pourrait aussi exprimer des propositions dont la forme est [pour y% de x, Fx], où y est un nombre entre 50 et 100, ou bien encore d'autres propositions statistiques, par exemple des propositions à propos de la fréquence relative des propriétés.

---

<sup>3</sup> Mais voir Barsalou 1990 et la section 4 de ce chapitre.

- Une *théorie* exprime des propositions qui sont génériques, causales, nomologiques et/ou fonctionnelles, c'est-à-dire des propositions de la forme [génériquement, Fx], [p parce que q], [pour tout x, Fx], et [p pour que q]. Par exemple, un concept théorique d'oiseaux pourrait exprimer les propositions [les oiseaux chantent], [les oiseaux ont des ailes pour voler]...

Les différences sémantiques entre ces trois entités théoriques sont compatibles avec la thèse selon laquelle la plupart des catégories sont représentées par plusieurs concepts qui appartiennent à différents types (voir la troisième interprétation de l'hypothèse de l'hétérogénéité, chapitre 2, section 2.3.). Pour la plupart des catégories au moins, il est possible d'encoder des connaissances sur les propriétés de certains de leurs membres, des connaissances sur les propriétés typiques de leurs membres et des connaissances génériques ou causales à propos de leurs membres.

## 2.2. Les propriétés fonctionnelles : ressemblance, prototypes et exemplaires

Les modèles des processus cognitifs associés au paradigme postulant des prototypes (PP) et au paradigme postulant des exemplaires (PE) sont *fondés sur la ressemblance* (« similarity-based ») : les processus cognitifs dans lesquels les exemplaires ou les prototypes sont supposés figurer sont censés reposer sur le calcul de la ressemblance entre une représentation et le prototype ou les exemplaires (sur la question du rôle de la ressemblance dans les processus cognitifs, voir le numéro spécial de *Cognition*, 1998, 65 et Hahn et Ramscar 2001). Selon ces deux approches, quand un individu s'engage dans une tâche conceptuelle, par exemple quand il catégorise, quand il raisonne déductivement ou inductivement, quand il résout un problème, quand il fait une analogie..., il calcule la ressemblance entre une représentation et un prototype ou un ensemble d'exemplaires.

Pour rendre cette idée un peu moins abstraite, je recense ci-dessous certaines explications typiques de nos compétences cognitives avancées par les partisans du PP et par les partisans du PE. Ces explications illustrent comment ces deux entités théoriques sont associées à des processus fondés sur la ressemblance. Elles reposent toutes sur une même idée. On trouve des phénomènes de typicalité dans la plupart de nos performances cognitives supérieures : nos performances sont affectées par le degré selon lequel un

objet (ou une sous-classe d'objets) est typique de sa catégorie. Les psychologues supposent que les théories des concepts doivent être à même d'expliquer ces phénomènes. Selon le PP et le PE, les phénomènes de typicalité proviennent de processus impliquant un calcul de ressemblance avec des prototypes ou des exemplaires.<sup>4</sup>

- *Catégorisation*<sup>5</sup>

Tous les modèles de la catégorisation qui s'inspirent du PP et du PE reposent sur le calcul de la ressemblance entre le prototype ou l'ensemble d'exemplaires et la cible (voir, par exemple, Hampton 1992, 2001, Medin et Schaffer 1978, Nosofsky 1986, 1988, Ross et Makin 1999 et le chapitre 5 du présent travail). En substance, le système cognitif est supposé calculer cette ressemblance conformément à une mesure particulière de la ressemblance (voir ci-dessous). Les cibles sont catégorisées dans une catégorie donnée seulement si leur ressemblance avec la représentation de cette catégorie est supérieure à un seuil de catégorisation.

Ces modèles ont été en partie inventés pour expliquer certains aspects bien connus des performances humaines de catégorisation, à savoir les phénomènes de typicalité dans les tâches de catégorisation.<sup>6</sup> Les sujets sont souvent en mesure de répondre à la question « dans quelle mesure cet objet, par exemple un pigeon, est typique de la catégorie des oiseaux ? ». Le degré de typicalité qui est attribué à des objets est corrélé à de nombreuses propriétés des performances de catégorisation des sujets : en particulier, les objets typiques (ou les sous-classes<sup>7</sup>) sont catégorisés plus rapidement et avec moins d'erreurs que les objets atypiques (Smith et al. 1974, Rosch et Mervis 1975, Rosch 1978, voir aussi le chapitre 5, section 4.1.). Les modèles de la catégorisation qui font appel à

---

<sup>4</sup> Voir Armstrong et al. 1983 et Barsalou 1985 contre ces explications. Gopnik et Meltzoff proposent une explication de la typicalité qui est compatible avec le paradigme postulant des théories (1997, 164). Pour une discussion des multiples facteurs qui affectent le degré de typicalité d'un objet, voir Murphy 2002. Les jugements de typicalité sont influencés surtout par la ressemblance entre la représentation de la catégorie (prototype ou ensemble d'exemplaires) et l'objet. Cependant, ils peuvent aussi être influencés par la fréquence de l'objet et par la familiarité avec l'objet (Armstrong et al. 1983, Barsalou 1985, Nosofsky 1988) et, dans une moindre mesure, par la comparaison avec un membre idéal de la catégorie (Barsalou 1985). Si la fréquence est contrôlée, la typicalité dépend presque entièrement de la ressemblance entre l'objet et la représentation de la catégorie.

<sup>5</sup> Voir le chapitre 5 pour une discussion plus approfondie de la catégorisation.

<sup>6</sup> La découverte de ces phénomènes fut à l'origine du rejet du paradigme classique des concepts (PC) au profit du PP.

<sup>7</sup> Par exemple, la classe des pigeons.

des prototypes ou à des exemplaires expliquent ces phénomènes de typicalité : les cibles les plus typiques ressemblent davantage au prototype ou aux exemplaires en sorte que le seuil de catégorisation est atteint plus rapidement et avec une plus grande probabilité (voir le chapitre 5, section 4.).

- *Le raisonnement inductif*

Considérons le modèle de l'induction fondée sur les catégories (« category-based induction model ») proposé par Osherson et al. (1990, 1991). Le but de ce modèle est d'expliquer la certitude qui est attribuée à la conclusion d'un raisonnement inductif.

Les membres de la catégorie  $A_1$  ont la propriété P

Les membres de la catégorie  $A_2$  ont la propriété P

Les membres de la catégorie  $A_3$  ont la propriété P

Quelle est la probabilité que les membres de la catégorie B aient la propriété P

De nombreux facteurs affectent la force des conclusions inductives (par exemple, la ressemblance entre les catégories mentionnées dans les prémisses et la catégorie mentionnée dans la conclusion, la connaissance à propos de la variabilité de la catégorie etc.<sup>8</sup>). Le modèle de l'induction fondée sur les catégories soutient qu'une inférence est d'autant plus forte que la catégorie de la prémisse est *typique* de sa catégorie superordonnée (sur la notion de catégorie superordonnée, voir le chapitre 2, note 15).<sup>9</sup> Osherson et collègues soutiennent que la force d'une conclusion inductive repose sur un processus de comparaison (« matching process ») dans les quels figurent des prototypes (ou des ensembles d'exemplaires<sup>10</sup>). Pour déterminer la typicalité, on doit calculer la ressemblance entre une instance et un prototype (ou un ensemble d'exemplaires) ou entre le prototype de la catégorie subordonnée et le prototype de la catégorie superordonnée (ou ensembles d'exemplaires). En d'autres mots, dans ce modèle, les processus cognitifs qui réalisent notre compétence inductive incluent une comparaison entre un prototype ou

---

<sup>8</sup> L'effet de ces variables peut aussi être expliqué au moyen des prototypes ou des exemplaires.

<sup>9</sup> On trouve de nombreux effets de typicalité dans nos raisonnements inductifs (Smith 1989, Osherson et al. 1990, Sloman 1993).

<sup>10</sup> Osherson et al. formulent leur hypothèse en termes de prototypes, mais l'on pourrait la reformuler en termes d'exemplaires.

un ensemble d'exemplaires et d'autres représentations (qui peuvent elles-mêmes être des prototypes ou des ensembles d'exemplaires) et un calcul de ressemblance entre ces représentations.

Ce modèle explique de nombreuses données empiriques. Les deux inférences inductives ci-dessous illustrent les données qui viennent à l'appui de ce modèle :

Les moineaux peuvent attraper la maladie appelée Ermatropieus

Les autruches peuvent attraper la maladie appelée Ermatropieus

Les pingouins peuvent attraper la maladie appelée Ermatropieus

Les autruches peuvent attraper la maladie appelée Ermatropieus

Clairement, la conclusion de la seconde inférence est plus faible que la conclusion de la première. Dans le modèle de l'induction fondée sur les catégories, cela résulte du fait que les pingouins sont des oiseaux moins typiques que les moineaux.

- *Le raisonnement déductif*

Les individus ne sont pas de très bons logiciens, comme le montrent nos difficultés à appliquer la règle du *modus tollens* (Wason et Johnson-Laird 1972, Cosmides 1989, Cosmides et Tooby 1992, mais voir Sperber et al. 1995, Girotto et al. 2001, Sperber et Girotto 2002). Hampton (1982, 1988) a trouvé d'autres déviations par rapport aux règles de déduction de la logique du premier ordre. L'une de ces déviations est le phénomène d'intransitivité<sup>11</sup> : certains sujets sont prêts à dire que les x sont y et que les y sont z, tout en niant que les x soient z. Par exemple, dans une expérience de Hampton (1982), « people were willing to accept that chairs are a type of furniture and that car seats are a

---

<sup>11</sup> Voir aussi le phénomène de surextension (« overextension ») dans Hampton 1988 et le chapitre 6 du présent travail. Sloman (1998) a aussi montré que les individus acceptent davantage la validité d'un argument déductif quand les catégories mentionnées dans les prémisses sont semblables à la catégorie de la conclusion que quand elles ne le sont pas.

type of chair, but would then deny that car seats are a type of furniture » (Hampton 1997b, 143).<sup>12</sup>

Cette déviation peut être vue comme un effet de typicalité (Hampton 1982, 1988, 1997b) : la ressemblance entre le prototype de chaise et le prototype de meuble et la ressemblance entre le prototype de chaise et le prototype de siège de voiture sont supérieures au seuil de catégorisation, tandis que la ressemblance entre le prototype de siège de voiture et le prototype de meuble ne l'est pas (Shafir et al. 1990 et Sloman 1998 présentent d'autres erreurs de raisonnement qui peuvent être expliquées au moyen d'un calcul de ressemblance avec des prototypes ou des ensembles d'exemplaires). Les explications du raisonnement déductif par le PP et le PE supposent donc respectivement que les prototypes et les exemplaires figurent dans des processus qui reposent sur la ressemblance.

- *Les raisonnements probabilistes et les choix en situation d'incertitude*

Le programme de recherche « Heuristiques et biais » (« Heuristics and Biases »), lancé par Tversky et Kahneman, a soutenu que les jugements de probabilité et les choix en situation d'incertitude violent souvent les normes de la théorie bayésienne. Ces violations sont expliquées au moyen d'heuristiques qui, dans de nombreuses circonstances, biaisent le raisonnement des individus. L'une de ces heuristiques est appelée *le biais de la représentativité* (« the representativeness bias ») : les individus utilisent souvent la typicalité d'un objet pour faire des jugements de probabilité à son propos, alors même que de tels jugements violent les principes de base de la théorie des probabilités (Kahneman et Tversky 1972, Tversky et Kahneman 1983, mais voir Gigerenzer 1991, 1996).

Ce biais est manifeste dans la fameuse *erreur de la conjonction* (Tversky et Kahneman 1983, mais voir Hertwig et Gigerenzer 1999) : dans le cas fameux de Linda, les sujets jugent qu'il est plus probable qu'une étudiante qui est une activiste politique devienne une employée de banque féministe plutôt qu'une employée de banque – violant ainsi la règle de la conjonction de la théorie des probabilités. Le biais de la

---

<sup>12</sup> Ce phénomène peut être expliqué de plusieurs manières. Par exemple, la sémantique de « un type de » peut expliquer les performances des sujets. Une explication pragmatique de ces phénomènes est aussi possible.

représentativité fournit une explication naturelle de ce phénomène : le prototype (ou un ensemble d'exemplaires) d'une étudiante activiste ressemble davantage au prototype (ou à un ensemble d'exemplaires) d'une employée de banque féministe qu'au prototype (ou qu'à un ensemble d'exemplaires) d'une employée de banque. En conséquence, les sujets jugent qu'une étudiante activiste a plus de chances de devenir une employée de banque féministe qu'une employée de banque (voir Smith et Osherson 1989 pour une modélisation simple de ce phénomène en accord avec le PP).

Le biais de la représentativité est un biais important dans les raisonnements de probabilité (Kahneman et Tversky 1972, Tversky et Kahneman 1983, Smith et Osherson 1989, mais voir Gigerenzer 1996). En plus de la violation de la règle de conjonction, il explique aussi la négligence bien connue du taux de base dans les jugements de probabilité (Kahneman et Tversky 1972, mais voir Gigerenzer et Hoffrage 1995). Smith et Osherson ont aussi proposé que les modèles du choix fondés sur la ressemblance pourraient expliquer d'autres données trouvées par Tversky et Kahneman, par exemple le biais du status quo (Smith et Osherson 1989, 71-72).

En résumé, les deux entités théoriques examinées pour l'instant, à savoir les exemplaires et les prototypes, sont supposées figurer dans des processus cognitifs qui reposent sur la ressemblance. La ressemblance entre des objets et des prototypes ou des ensembles d'exemplaires est supposée expliquer les propriétés de nos processus cognitifs, en particuliers, les phénomènes de typicalité. Ainsi Hampton écrit-il (2001, 20) :

« Similarity effects are pervasive in people's attempts to reason logically, and a very simple explanation of this finding is that our conceptual system is heavily dependent on similarity-based conceptual processes. »

### 2.3. Les propriétés fonctionnelles : calculs additif et multiplicatif de la ressemblance

Les prototypes et les exemplaires sont supposés figurer dans des processus cognitifs qui reposent sur différentes mesures de la ressemblance.

- Les prototypes sont souvent associés à des mesures *linéaires* de la ressemblance (voir Hampton 1993 et Ross et Makin 1999), tandis que les exemplaires sont souvent

associés à des mesures *non linéaires* (voir Medin et Schaffer 1978). Dans les deux cas, le système cognitif concerné calcule la ressemblance entre un concept et un objet en comparant les propriétés qui sont représentées par le concept (par le prototype ou par les exemplaires) et celles qui sont attribuées à l'objet. La ressemblance entre la cible et l'exemplaire est une fonction du nombre de propriétés communes. Dans le cas d'une mesure non linéaire, son accélération est positive.

- Les mesures linéaires sont *additives*. Dans une mesure additive de la ressemblance, on *ajoute* les propriétés communes aux deux représentations pour calculer leur ressemblance (voir, par exemple, la règle du contraste proposée par Tversky 1977). Je présente ci-dessous un exemple de mesure additive de la ressemblance (voir Hampton 1993 et Ross et Makin 1999) :

$S(A,t)$  est la mesure de la ressemblance de la cible  $t$  par rapport au prototype de la catégorie  $A$ ;

$$S(A,t) = \sum_{i=1}^n (w_i \times v_{it})$$

où  $w_i$  est le poids de la  $i^{\text{e}}$  propriété du prototype (*rond, rouge, utilisé en salade* etc.), et  $v_{it}$  est fonction du degré auquel la cible possède la  $i^{\text{e}}$  propriété;

les poids  $w_i$  représentent la fréquence de la possession de la  $i^{\text{e}}$  propriété par les membres de la catégorie dénotée.

- Les mesures non linéaires sont souvent *multiplicatives*.<sup>13</sup> Dans le cas d'une mesure multiplicative, on multiplie les valeurs qui expriment les degrés selon lesquels chaque propriété est commune à la cible et au concept (Ross et Makin 1999, Murphy 2002). Par exemple, on pourrait proposer que la mesure multiplicative de la ressemblance est la suivante (voir Medin et Schaffer 1978 et Ross et Makin 1999) :

$S(B,t)$  est la ressemblance de la cible par rapport à l'ensemble  $B$  d'exemplaires;

$$S(B,t) = \sum_{b \in B} S(b,t);$$

---

<sup>13</sup> Certaines théories des exemplaires utilisent d'autres mesures non linéaires (par exemple, Hintzman 1986).

$$S(b,t) = \prod_{i=1}^n (u_i \times v_{it})$$

où  $u_i$  est le poids de la  $i^e$  propriété de  $b$  et  $v_{it}$  est fonction du degré auquel la cible possède la  $i^e$  propriété;

si les poids  $u_i$  sont identiques, on obtient  $S(b,t) = \prod_{i=1}^n (v_{it})$ ;

souvent,  $v_{it} = 1$  quand la cible possède la  $i^e$  propriété et  $v_{it} = m$  ( $0 < m < 1$ ) quand elle ne la possède pas

- Je présente un exemple simplifié pour illustrer ces deux types de mesures (voir Ross et Makin 1999, 210-211). Supposons que la catégorie des pommes est dénotée par le prototype (fictif) suivant :

<b>POMME<sub>pr</sub></b>	
<i>Propriété</i>	<i>Poids</i>
Rouge	1
Rond	1
Lisse	0.5
Se mange crue	0.5

**Schéma 4.1.**

Supposons aussi que la catégorie des pommes est dénotée par les deux exemplaires (fictifs) suivants :

<b>POMME<sub>ex1</sub></b>	
<i>Propriété</i>	<i>Degré de possession</i>
Rouge	1
Rond	1
Lisse	1
Se mange crue	1

**Schéma 4.2.**

POMME <sub>ex2</sub>	
<i>Propriété</i>	<i>Degré de possession</i>
Rouge	1
Rond	1
Lisse	0
Se mange crue	0

**Schéma 4.3.**

Le prototype POMME<sub>pr</sub> est abstrait à partir des individus représentés par les deux exemplaires POMME<sub>ex1</sub> et POMME<sub>ex2</sub>.

Considérons maintenant une nouvelle pomme :

Cible	
<i>Propriété</i>	<i>Degré de possession</i>
Rouge	0
Rond	1
Lisse	1
Se mange crue	1

**Schéma 4.4.**

Dans les exemplaires, les propriétés *rouge*, *rond*, *lisse* et *se mange crue* sont pondérées de la même manière (voir la variable  $u_i$  ci-dessus) : chaque propriété est également importante pour être catégorisée comme une pomme quand on compare la cible à des exemplaires. La variable  $v_{it}$  est égale à 1 si la cible possède la propriété ou à 0 si elle ne la possède pas. Quand nous calculons le degré de ressemblance entre la cible et le prototype de la catégorie de pomme d'une part et le degré de ressemblance entre la cible et l'ensemble des exemplaires de l'autre, nous obtenons deux degrés de ressemblance différents, alors même que le prototype est abstrait à partir des deux individus représentés par les exemplaires. En effet, quand on utilise la mesure additive, on obtient:

$$S(\text{POMME}_{\text{pr}}, \text{cible}) = m + 1 + 0.5 + 0.5 = 2 + m$$

où  $m$  varie entre 0 et 1.

Quand on utilise la mesure multiplicative, on obtient:

$$S(\{\text{POMME}_{\text{ex1}}, \text{POMME}_{\text{ex2}}\}, \text{cible}) = S(\text{POMME}_{\text{ex1}}, \text{cible}) + S(\text{POMME}_{\text{ex2}}, \text{cible})$$

$$S(\text{POMME}_{\text{ex1}}, \text{cible}) = m \times 1 \times 1 \times 1$$

$$S(\text{POMME}_{\text{ex2}}, \text{cible}) = m \times 1 \times m \times m$$

$$S(\{\text{POMME}_{\text{ex1}}, \text{POMME}_{\text{ex2}}\}, \text{cible}) = m + m^3$$

où  $m$  varie entre 0 et 1.

- Il est en principe possible de combiner les prototypes avec une mesure non linéaire de la ressemblance et les exemplaires avec une mesure linéaire de la ressemblance. Toutefois, les mesures non linéaires de la ressemblance s'accordent mieux avec le PE que les mesures linéaires. En effet, si la ressemblance entre une cible et un ensemble  $S$  d'exemplaires repose sur une mesure non linéaire, une cible peut ressembler davantage à  $S$  quand elle ressemble beaucoup à un unique exemplaire et peu aux autres, par exemple si elle partage  $n$  propriétés avec un exemplaire et 0 avec  $n-1$  autres exemplaires, que quand elle ressemble un peu à chaque exemplaire, par exemple si elle a en commun une seule propriété avec  $n$  exemplaires. Cela introduit donc un biais en faveur de la ressemblance entre la cible et des instances spécifiques de la catégorie.<sup>14</sup> Ce n'est pas le cas si la ressemblance repose sur une mesure linéaire. Comparons, par exemple, la mesure proposée par Medin et Schaffer (voir ci-dessus) et la mesure additive suivante :

$S(B,t)$  est la ressemblance de la cible par rapport à l'ensemble  $B$  d'exemplaires;

---

<sup>14</sup> Cela permet d'expliquer *l'effet exemplaire* (« the exemplar effect »), c'est-à-dire le fait qu'un objet atypique est catégorisé plus rapidement et avec moins d'erreurs qu'un membre typique si cet objet atypique est très semblable à un membre connu de la catégorie (Brooks 1978, Medin et Schaffer 1978, Nosofsky 1986, Whittlesea 1987, et le chapitre 4, section 4.1.)

$$S(B,t) = \sum_{b \in B} S(b,t);$$

$$S(b,t) = \sum_{i=1}^n (v_{it})$$

où  $v_{it} = 1$  quand la cible possède la  $i^e$  propriété et une constante  $m$  ( $0 < m < 1$ ) quand elle ne la possède pas.

Supposons d'abord qu'une cible possède en commun 1 propriété avec 5 objets parmi les 5 propriétés qui caractérisent ces objets. Etant donné la mesure additive de la ressemblance, la ressemblance de la cible avec ces 5 objets est alors égale à  $5 + 20m$ . Considérons maintenant le cas où la cible a 5 propriétés en commun avec un objet et 0 avec 4 autres parmi les 5 propriétés qui caractérisent ces objets. La ressemblance de la cible avec ces 5 objets est à nouveau égale à  $5 + 20m$ . Appliquons maintenant la mesure multiplicative proposée par Medin et Schaffer. Dans le premier cas, la ressemblance est de  $5m^4$  tandis que dans le second cas, la ressemblance est de  $1 + 4m^5$ . On voit donc qu'avec la mesure multiplicative, la ressemblance avec un exemplaire particulier parmi tous les exemplaires joue un rôle très important. Or, c'est précisément pour expliquer ce fait que la notion d'exemplaire a été introduite (Brooks 1978, Medin et Schaffer 1978) : l'idée était d'expliquer pourquoi nos catégorisations reposent parfois sur la ressemblance entre la cible et un objet particulier que nous avons antérieurement rencontré.

Inversement, les mesures linéaires de la ressemblance s'accordent mieux avec le PP que les mesures non linéaires. Car la ressemblance entre une cible et un objet particulier n'est pas valorisée si la mesure est linéaire (voir ci-dessus). Or le PP repose sur l'idée que nos concepts ne représentent pas des objets particuliers : nos concepts sont des connaissances générales. En conséquence, le calcul de ressemblance associé au PE met en valeur la ressemblance avec des objets particuliers de la catégorie, contrairement au calcul de ressemblance associé au PP.

#### *2.4. Les propriétés fonctionnelles: théories et processus fondés sur les règles*

Selon le PT, les théories ne figurent pas dans des processus cognitifs fondés sur la ressemblance. Car selon la plupart des avocats du PT, les modèles des processus cognitifs qui reposent sur la ressemblance sont incapables de caractériser les processus dans lesquels figurent des connaissances théoriques (mais voir Heit 2001). Au contraire, les

théories sont supposées figurer dans un autre type de processus, habituellement appelé « *processus cognitifs fondés sur des règles* » (« rule-based processes »).

Cette idée doit être explicitée.<sup>15</sup> La notion de règle est un peu vague et a été utilisée dans plusieurs débats théoriques, particulièrement dans le débat entre les modèles classiques et les modèles connexionnistes de l'esprit (Fodor et Pylyshyn 1988, Pinker et Prince 1988, 1999, Andler 1990, Davies 1991, Clark 1993, Marcus 2001) et dans les débats entre diverses conceptions de nos compétences cognitives (Pinker et Prince 1988, 1999, Smith et al. 1992, Medin et al. 1993, voir aussi le numéro spécial de *Cognition* 1998, 65).<sup>16</sup>

En un sens, un processus fondé sur la ressemblance repose simplement sur un type particulier de règles – à savoir sur une règle qui fait référence à la ressemblance entre représentations. Par exemple, en un sens, les modèles de la catégorisation fondés sur la ressemblance admettent la règle « si la cible est suffisamment semblable au prototype de la catégorie – ou à l'ensemble d'exemplaires –, la catégoriser dans la catégorie X ». En outre, calculer la ressemblance entre deux représentations, c'est appliquer une règle (par exemple, la règle du contraste de Tversky 1977). En ce sens, l'opposition entre les modèles fondés sur la ressemblance et les modèles fondés sur les règles est inadéquate.

Cependant, en un autre sens, il y a une claire opposition entre les règles et la ressemblance. Un processus repose sur la ressemblance si le traitement d'une représentation est gouverné par la ressemblance de cette représentation avec d'autres représentations. Un processus repose sur des règles si le traitement d'une représentation n'est pas gouverné par la ressemblance de cette représentation avec d'autres représentations, mais par l'appartenance de cette représentation à une certaine classe de représentations. Considérons l'exemple simpliste suivante. Supposons que la règle linguistique ci-dessous fasse partie de notre grammaire :

NP + VP → S.

---

<sup>15</sup> Je voudrais remercier ma commentatrice à la SPP 2003, Michelle Montague, pour avoir souligné ce point.

<sup>16</sup> Ces deux débats sont liés, bien que d'une manière complexe.

Elle s'applique à la classe des représentations des clauses nominales. Autrement dit, la manière dont les représentations des clauses nominales sont traitées n'est pas déterminée par leur ressemblance avec d'autres représentations, mais par le fait qu'elles appartiennent à la classe des représentations de clauses nominales. Généralement, les mécanismes cognitifs qui reposent sur des théories sont supposés traiter les représentations *indépendamment de leur ressemblance avec d'autres représentations, mais en fonction de leur appartenance à une classe de représentations*.<sup>17</sup>

Pourquoi les théories sont-elles supposées figurer dans des processus cognitifs qui dépendent de règles ? Pour deux raisons assez différentes. *Tout d'abord, les théories fournissent des règles qui sont appliquées dans les processus cognitifs*. Par exemple, selon la théorie de l'essentialisme psychologique (voir le chapitre 3, section 2.1., et le chapitre 7, section 1.), les concepts théoriques d'animaux fournissent des règles d'appartenance à la catégorie : le concept théorique de chien fournit la règle « les chiens sont nécessairement engendrés par des chiens » et cette règle est supposée être appliquée par le processus cognitif de catégorisation. Selon cette règle, Médor – qui a été engendré par des loups – est jugé ne pas être un chien, quoiqu'il ressemble à un chien. Médor n'est pas catégorisé comme un chien parce que la représentation de Médor n'appartient pas à la classe des représentations d'individus qui ont été engendrés par des chiens. Le processus de catégorisation est donc fondé sur une règle. *Deuxièmement, les concepts théoriques sont des ensembles de connaissances causales, nomologiques et fonctionnelles*. Et ce type de connaissances n'est pas utilisé dans des processus de comparaison qui calculent la ressemblance entre deux représentations – en déterminant le nombre de propriétés représentées à la fois par les deux représentations, mais par des raisonnements causaux ou dans des inférences à la meilleure explication. Ces types de raisonnement sont supposés obéir aux règles de la logique déductive et de la théorie des probabilités. En résumé, les concepts théoriques sont supposés figurer dans des processus fondés sur des règles pour

---

<sup>17</sup> Soulignons quelques fausses pistes. Tout d'abord, les règles n'ont pas à être explicites. La distinction implicite / explicite est orthogonale au problème examiné (contre Sloman et Rips 1998). Ensuite, les processus fondés sur la ressemblance ne sont pas essentiellement liés aux modèles connexionnistes de la cognition. Une architecture classique peut implémenter les processus fondés sur la ressemblance. Cependant, si les réseaux neuronaux sont vus comme des modèles cognitifs – et non comme des implémentations (voir Fodor et Pylyshyn 1988 et Ramsey et al. 1989 sur cette distinction), le connexionnisme est essentiellement un modèle de la cognition fondé sur la ressemblance.

deux raisons : les processus cognitifs sont supposés obéir aux règles du raisonnement (logique déductive et théorie des probabilités) et les concepts théoriques eux-mêmes fournissent des règles qui sont appliquées dans ces processus.<sup>18</sup>

Quels sont donc ces processus fondés sur des règles ? Quand l'on cherche à répondre à cette question, on est frappé par le contraste entre le PP et le PE d'un côté et le PT de l'autre. Le PP et le PE sont d'habitude associés à des modèles précis des processus cognitifs (voir, par exemple, Nosofsky 1986, Osherson et Smith 1989, Smith et al. 1988). Au contraire, le PT n'a pas développé de modèles précis. Les partisans du PT se satisfont souvent de rejeter les modèles fondés sur la ressemblance et parlent vaguement de modèles cognitifs fondés sur les règles (Heit 2001 fait le même constat).

Cependant, comme je l'ai fait pour le PP et le PE, je présente ci-dessous des explications typiques des processus cognitifs qui font appel aux concepts théoriques. Ces exemples illustrent la manière dont les théories sont supposées figurer dans les processus cognitifs fondés sur des règles.

- *Catégorisation*<sup>19</sup>

Dans leur article fondateur, Murphy et Medin (1985/2004) imaginent la catégorisation suivante. Nous sommes à une soirée. Un invité plonge tout habillé dans la piscine. Nous concluons qu'il est ivre. En d'autres mots, nous catégorisons le plongeur comme étant ivre. Cependant, il est peu probable que notre concept PERSONNE IVRE encode la moindre connaissance à propos des plongeurs tout habillé dans les piscines. En conséquence, il est peu probable que notre catégorisation repose sur la ressemblance entre le comportement de l'invité et notre concept d'individu ivre : nous ne comparons pas notre concept d'individu ivre à la cible. Au lieu de cela, Murphy et Medin proposent que nous tirons cette conclusion, parce que *c'est la meilleure explication* du comportement de cet invité. Le concept théorique PERSONNE IVRE est supposé servir de prémisse à ce raisonnement causal. Les avocats du PT soutiennent ainsi souvent que

---

<sup>18</sup> Les concepts théoriques sont supposés figurer dans des processus cognitifs semblables aux explications scientifiques. Et les théories scientifiques figurent dans des explications qui reposent sur des règles et non dans des processus de comparaison.

<sup>19</sup> Voir le chapitre 5, section 3.2..

nos décisions de catégorisation reposent sur des inférences à la meilleure explication – qui reposent elles-mêmes sur les règles de la théorie des probabilités (voir Rips 1989 et le chapitre 5, sections 2. et 3.).<sup>20</sup>

Les décisions de catégorisation peuvent aussi être fondées sur des règles dérivées de nos théories. Les avocats de la théorie de l'essentialisme psychologique admettent souvent que les performances de catégorisation sont gouvernées par la règle « si un objet t est un x, aucune modification de t ne modifie son appartenance catégorielle – à moins que cette modification n'affecte l'essence de t ». Par exemple, le concept théorique CHAT pourrait contenir la règle suivante : si un objet t est un chat, aucune modification de t ne modifie son appartenance catégorielle – à moins que cette modification n'affecte l'essence des chats. Ils prétendent expliquer ainsi le fait que les individus ne modifient pas leurs jugements de catégorisation quand l'apparence d'un animal est modifiée (Keil 1989, Gelman et Hirschfeld 1999, voir le chapitre 7, section 1.).

Ces deux idées sont assez différentes. Dans un cas, les règles sont supposées être les règles générales du raisonnement (principes du raisonnement probabiliste, règles de déduction etc.), tandis que dans l'autre, les règles sont des principes particuliers qui sont spécifiés par nos concepts théoriques. Mais, dans les deux cas, les processus cognitifs dans lesquels sont censés figurer les concepts théoriques ne reposent pas sur un calcul de ressemblance.

- *Formation de catégories*<sup>21</sup>

La formation des catégories que les concepts théoriques sont censés dénoter a été décrite comme *une inférence à la meilleure explication* (par exemple, Schank et al. 1986, mentionné dans Rips 1989, voir le chapitre 5, section 3.2.).<sup>22</sup> L'idée est la suivante. Selon les avocats du PT, dans des conditions écologiquement valides (par opposition aux conditions expérimentales qui utilisent des catégories artificielles définies par une ou deux propriétés arbitraires), former une catégorie, c'est *fournir* une *raison* pour grouper

---

<sup>20</sup> Heit écrit (2001, 156) : « A great deal of work (...) has contrasted the similarity-based approach to categorization and the knowledge-based approach. The similarity-based approach highlights the use of similarity to observed category members, and the knowledge-based approach emphasizes other processes and use of theoretical knowledge. »

<sup>21</sup> Voir le chapitre 5, section 3.2..

<sup>22</sup> Rips écrit (1989, 53) : « categorization and category learning are special cases of inference to the best explanation. »

des objets dans une même classe. Les concepts théoriques sont les produits de ces inférences à la meilleure explication.

Selon les partisans du PT, la plupart de nos catégories réelles reposent sur ce type d'explications. Murphy et Medin (1985/2004) proposent l'exemple suivant.

« Pourquoi certains objets sont-ils regroupés pour former une catégorie ? (...) Un exemple quelque peu inhabituel, mais cependant utile, vient d'une vieille énigme de l'érudition biblique, les règles alimentaires associées avec les abominations du Lévitique, qui produisent les catégories *animaux purs* et *animaux impurs*. Pourquoi les chameaux, les autruches, les crocodiles, les souris, les requins, et les anguilles devraient-ils être déclarés impurs, tandis que les gazelles, les grenouilles, la plupart des poissons, les sauterelles et certains criquets devraient être déclarés purs? Que peuvent bien avoir de commun les caméléons, les taupes et les crocodiles en sorte qu'ils soient groupés les uns avec les autres ? C'est-à-dire, qu'est-ce qui rend les catégories des animaux purs et impurs sensées ou cohérentes ? »

Ils soulignent l'explication suivante :

« Selon nous, la spéculation la plus convaincante concernant cette règle de classification a été développée dans le livre intrigant de Mary Douglas (1966), *Purity and Danger* : selon elle, il doit y avoir une corrélation entre le type d'habitat, la structure biologique, et la forme de locomotion. Les créatures aquatiques doivent avoir des nageoires et des écailles, et doivent nager ; les créatures terrestres doivent avoir quatre pattes et doivent sauter ou marcher ; les créatures aériennes doivent voler avec des ailes plumées. Toute classe de créatures non équipées pour le genre de locomotion appropriée à son élément est impure. Par exemple, les autruches seraient impures parce qu'elles ne volent pas. Les crocodiles sont impurs parce que leurs pattes de devant ressemblent à des mains, alors qu'ils marchent sur leurs quatre pattes. »

- *Raisonnement inductif*

Selon le PT, les inductions ne font pas appel à des relations de ressemblance entre représentations mentales, mais aux règles probabilistes appliquées aux connaissances théoriques encodées par nos concepts. Par exemple, dans certaines tâches inductives, les Itza Maya, qui vivent depuis des générations dans les forêts du Guatemala, raisonnent différemment des Occidentaux. Selon Coley et al. (1997, 1999, Atran 1999), les Occidentaux jugent que les arguments inductifs dont les prémisses mentionnent des catégories différentes (voir (2) ci-dessous) produisent des conclusions plus fortes que les arguments dont les prémisses mentionnent des catégories semblables (voir (1) ci-dessous).

(1) Jaguars have a disease Y  
Leopards have a disease Y  
All mammals have a disease Y.

(2) Jaguars have a disease Y  
Mice have a disease Y  
All mammals have a disease Y.

Les Itza Maya ont souvent des jugements opposés. Coley et collègues proposent que les connaissances biologiques que possèdent les Maya, peut-être leurs connaissances sur les relations écologiques entre animaux, expliquent pourquoi ils ne pensent pas que les raisonnements inductifs formés sur le modèle de (2) produisent systématiquement des conclusions plus fortes que les raisonnements formés sur le modèle de (1). Selon les avocats du PT, ce type de connaissances théoriques est encodé par des concepts théoriques.

Par ailleurs, nos raisonnements inductifs ne semblent pas toujours reposer sur la ressemblance entre représentations (comme le soutiennent Osherson et al. 1990 – voir ci-dessus). L'induction semble parfois obéir aux règles probabilistes. Smith et collègues (1992, 13-14) ont ainsi résumé les nombreuses données qui montrent que les individus

sont capables de suivre les règles de la théorie des probabilités lorsqu'ils raisonnent inductivement. Par exemple, nous utilisons la loi statistique des grands nombres.

### *2.5. Conclusion*

Les trois entités théoriques que nous avons considérées, les exemplaires, les théories et les prototypes, sont associées à des types de processus cognitifs différents. Les exemplaires et les prototypes sont supposés figurer dans des processus qui reposent sur des calculs de ressemblance entre représentations. Ils se distinguent par la nature de la ressemblance mise en jeu dans ces processus : les prototypes sont associés à une mesure linéaire de la ressemblance, tandis que les exemplaires sont associés à une mesure non linéaire, souvent multiplicative. Les théories sont supposées figurer dans des processus qui reposent sur des règles. Tout d'abord, les connaissances causales et nomologiques supposent des processus cognitifs qui obéissent aux règles du raisonnement probabiliste et déductif. Ensuite, les concepts théoriques fournissent des règles qui sont appliquées dans les processus cognitifs. En conséquence, prototypes, exemplaires et théories sont supposés figurer dans des processus cognitifs fort différents. Les propriétés fonctionnelles de ces entités théoriques sont donc différentes.

J'ai montré que les prototypes, les exemplaires et les théories avaient des propriétés sémantiques et fonctionnelles différentes. Ces propriétés sont essentielles pour évaluer la ressemblance entre types de représentations. En conséquence, le PT, le PE et le PP posent trois entités théoriques qui ont peu en commun (prémisse 3 de l'ANE). Cela établit la thèse conditionnelle suivante : si l'on peut montrer que ces trois entités théoriques sont réelles, c'est-à-dire si l'on peut montrer que nous avons bien des prototypes, des exemplaires et des théories, alors les concepts se divisent en types qui ont peu en commun.

## **3. La théorie pluraliste des processus cognitifs**

### *3.1. Le problème*

Le lecteur peut être surpris par la conclusion de la section précédente. Car j'ai soutenu dans le chapitre 2 que la nature de nos concepts ne varie pas d'une compétence cognitive

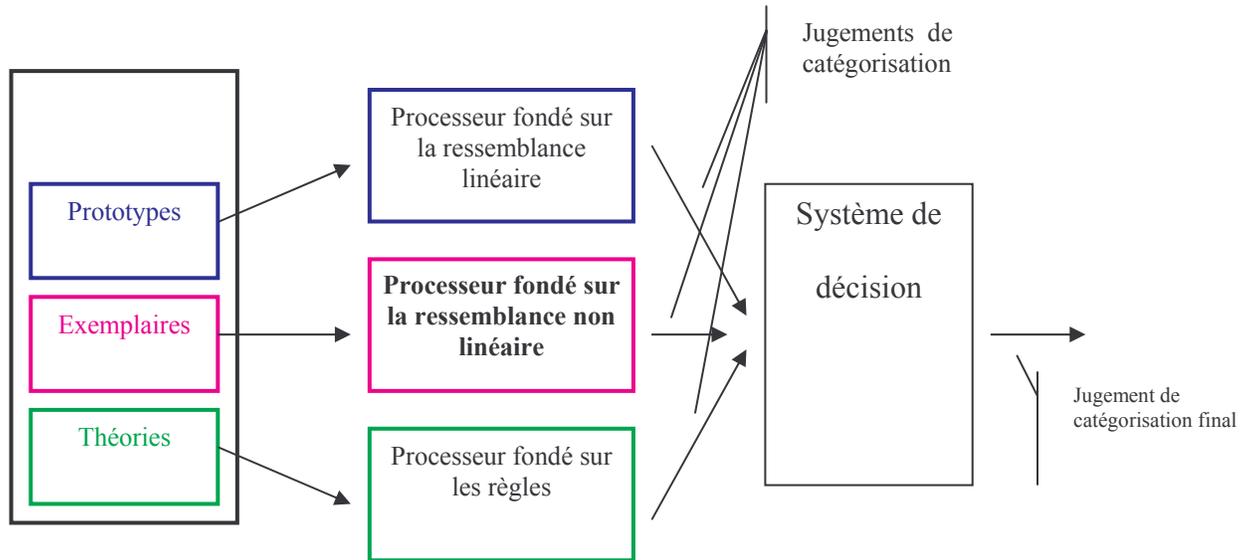
à l'autre (voir la première interprétation de l'hypothèse de l'hétérogénéité, section 2.3.) : il n'est pas vrai que nous ayons un type de concepts pour catégoriser, un autre type pour raisonner déductivement, un autre type encore pour raisonner inductivement... Or je soutiens maintenant que les entités théoriques postulées par les principaux paradigmes des concepts sont supposées figurer dans des types différents de processus, c'est-à-dire dans des processus fondés sur la ressemblance linéaire, dans des processus fondés sur la ressemblance non linéaire, et dans des processus fondés sur des règles. On pourrait peut-être objecter que cette conclusion est incompatible avec la thèse soutenue dans le chapitre 2. Comment puis-je en même temps nier que les concepts varient d'une compétence à l'autre et soutenir que les différents types de concepts figurent dans des processus cognitifs de type différent ?

### 3.2. *Compétence et processus*

Cependant, il n'y a là aucune incompatibilité. Car *les compétences* ne doivent pas être confondues avec des *processus*. Une compétence, par exemple la capacité de catégoriser ou de faire une induction, est réalisée par un ou plusieurs processus cognitifs (voir le chapitre 1, section 1.2.1. sur ces notions). Une compétence est la capacité de produire un output donné, étant donné un input spécifique. Un processus cognitif est une manière spécifique de produire cet output, étant donné cet input. *Or une même compétence, par exemple la capacité de catégoriser, peut être réalisée par plusieurs processus cognitifs indépendants les uns des autres.*<sup>23</sup> Par exemple, la catégorisation peut être réalisée par différents types de processus cognitifs (je défends précisément cette hypothèse dans le chapitre 5). Chaque processus est suffisant pour catégoriser un objet. Le schéma suivant illustre cette idée.

---

<sup>23</sup> On pourrait cependant faire l'objection suivante. Il n'y a pas de manière consensuelle d'individuer les compétences cognitives (et peut-être les capacités en général). En conséquence, on pourrait individuer nos compétences cognitives de telle sorte qu'il y ait une correspondance bijective entre elles et nos processus cognitifs. Les concepts varieraient alors d'une compétence à l'autre. Par exemple, on pourrait nier que nous ayons une compétence générale de catégorisation. A la place, nous pourrions avoir une compétence de catégorisation fondée sur la ressemblance linéaire, c'est-à-dire une capacité de classer des entités sur la base de la ressemblance linéaire, une compétence de catégorisation fondée sur la ressemblance non linéaire et une compétence de catégorisation fondée sur des règles. Chaque compétence serait réalisée par un processus cognitif unique. Cette objection a une certaine validité – étant donné qu'il n'y a pas de manière canonique d'individuer les compétences. Cependant, un prototype figurerait toujours dans plusieurs processus, par exemple dans la catégorisation fondée sur la ressemblance linéaire, dans le raisonnement inductif fondé sur la catégorisation linéaire...



**Schéma 4.5.**

### 3.3. Articulation des processus cognitifs

La manière dont ces processus sont articulés est bien sûr une question empirique (voir Smith et Sloman 1994, Smith et al. 1998, Hahn et Chater 1998 et le chapitre 5, section 4.5.). Cependant, on peut envisager certaines possibilités :

- Ces processus peuvent fonctionner en parallèle dans tous les cas. Par exemple, chaque fois que nous nous catégorisons un objet, tous les processus de catégorisation peuvent être déclenchés.
- Chaque processus peut être déclenché dans des circonstances particulières. Par exemple, le processus de catégorisation dans lequel les prototypes figurent peut être déclenché par un certain type d'inputs, tandis que le processus de catégorisation dans lequel les théories figurent peut être déclenché par un autre type d'inputs.
- Ils peuvent fonctionner en parallèle dans certaines circonstances, mais pas dans d'autres. Par exemple, on peut imaginer que, dans des raisonnements exprimés verbalement, le processus de catégorisation dans lequel les théories figurent est déclenché de manière préférentielle (Smith et Sloman 1994, voir aussi le chapitre 5, section 3.),

tandis que dans la catégorisation perceptuelle, tous les processus sont déclenchés en parallèle.

Si les processus fonctionnent en parallèle, leurs outputs doivent être *intégrés*, ou il doit exister une règle pour *choisir entre eux*. C'est une tâche cruciale quand ces outputs diffèrent (Allen et Brooks 1991, Smih et al. 1998, Hahn et Chater 1998, 224).<sup>24</sup> Par exemple, quand nous faisons une induction, le degré de certitude de la conclusion inductive peut être le résultat de l'intégration de différents processus. Elle pourrait aussi correspondre à l'output d'un seul processus, les outputs des autres processus ayant été négligés.

L'acceptation de l'hypothèse EN a conduit les psychologues à négliger cet objet d'étude empirique. Si l'on admet qu'il existe différents types de concepts qui figurent dans des processus cognitifs différents, la question de l'articulation de ces processus se pose aussitôt. Il faut espérer que cette question attirera l'attention des psychologues dans le futur.

### 3.4. La théorie pluraliste et les théories duelles de la cognition

*Cette position n'est nullement ad hoc.* Au contraire, elle ressemble par certains aspects aux *théorie duelles de la cognition*, qui sont aujourd'hui admises par un grand nombre de psychologues (« dual-process theories of cognition », voir Smith et al. 1992, Smith et Sloman 1994, Sloman 1996, Sloman et Rips 1998, Hahn et Chater 1998, Haidt 2001...). Ces théories postulent la thèse suivante (Haidt 2001, 819):

« two processing systems are often at work when a person makes judgments or solves problems. »

Ces processus fonctionnent souvent en parallèle et peuvent produire des outputs différents: par exemple, nos systèmes de catégorisation peuvent ne pas s'accorder sur

---

<sup>24</sup> Hahn et Chater écrivent (1998, 224): « (...) rules and similarity both have their respective roles, not just side by side, with similarity covering some domains and rules others, or « doubling up » in parallel (Sloman 1996), but in an *active interplay* within a single task. (...) This suggests a shift of emphasis in future research, from pitting rules against similarity toward experimental and computational investigation of the potential interplay of rules and similarity in cognition. »

l'appartenance de la cible à une catégorie. Plusieurs distinctions ont été proposées entre les deux types de processus postulés par les théories duelles. Certains psychologues ont distingué les processus conscients des processus inconscients, d'autres ont distingué les processus automatiques des processus volontaires etc. (voir Smith et al. 1998, Hahn et Chater 1998, et la figure 1 dans Haidt 2001, Stanovich et West 2000, 2003).

Certains psychologues croient que la distinction entre deux types de processus s'applique généralement à toute notre cognition. Ainsi le psychologue des émotions et de la moralité Jonathan Haidt soutient-il que nos jugements moraux peuvent reposer sur deux processus différents, l'un étant fondé sur des principes théoriques, l'autre sur nos intuitions spontanées et sur nos émotions (Haidt 2001, Greene et Haidt 2002). Les outputs de ces deux processus peuvent être différents (Greene et al. 2001). Le psychologue du raisonnement Steven Sloman soutient que nos inférences déductives reposent sur un système de raisonnement fondé sur des règles et sur un système fondé sur la ressemblance (Sloman 1996). Le second système fonctionne automatiquement, tandis que le premier est supposé avoir besoin d'un contrôle métacognitif et est moins fréquemment déclenché. Ces deux systèmes sont supposés produire souvent des outputs différents.

Comme les théories duelles, je soutiens que les compétences cognitives sont réalisées par plusieurs processus. *Mais, contrairement à ces théories duelles, je propose qu'il y a plus de deux systèmes: je soutiens en effet qu'il y a souvent au moins trois types de processus.* En outre, j'avance que ces systèmes impliquent différents types de représentations, à savoir des exemplaires, des prototypes et des théories.<sup>25</sup>

### 3.5. Conclusion

Toutes les compétences cognitives supérieures, par exemple la catégorisation, le raisonnement déductif, le raisonnement inductif, la capacité de faire des analogies etc., peuvent être réalisées par plusieurs processus. Le fait que les exemplaires, les prototypes et les théories figurent dans des processus cognitifs de nature différente n'implique pas que les représentations conceptuelles varient d'une compétence cognitive à l'autre. La

---

<sup>25</sup> Il se peut aussi que certaines compétences soient réalisées par un système qui intègre ces différents types de représentations. De fait, je soutiens que cela est le cas pour notre capacité de composer des concepts (voir le chapitre 6).

nature de ces processus et la manière dont ils sont articulés devraient attirer l'attention des psychologues.

## 4. L'objection de Barsalou

### 4.1. La thèse de Barsalou et ses arguments

Barsalou a avancé une thèse frappante (1990, 61)<sup>26</sup>:

« exemplars and abstracted representations in principle are informationally equivalent. Because of this equivalence, we can not determine whether people use exemplars or abstracted representations. »

Barsalou soutient donc que l'on ne peut distinguer expérimentalement le PP du PE. Il avance deux arguments :

- Certains modèles des concepts inspirés par le PP sont *informationnellement équivalents* à tous les modèles des concepts inspirés par le PE et certains modèles des concepts inspirés par le PE sont *informationnellement équivalents* à tous les modèles des concepts inspirés par le PP.

L'idée est la suivante. Deux modèles des concepts sont informationnellement équivalents s'ils supposent l'existence de représentations mentales qui encodent les mêmes connaissances. La thèse est donc que dans certains modèles des prototypes, les prototypes encodent les mêmes connaissances que des exemplaires pourraient encoder et vice-versa.

Barsalou ne nie pas qu'il y ait une différence sémantique entre la plupart des modèles des prototypes et la plupart des modèles des exemplaires. Il soutient tout d'abord que, dans *certaines* modèles des prototypes, les prototypes encodent toutes les connaissances à propos des individus que des exemplaires peuvent encoder (voir les exemples dans Barsalou 1990, figures 2.1., 2.2. et 2.3.).<sup>27</sup> Car ces prototypes peuvent encoder l'instanciation unique d'une propriété ou la coinstanciation unique de plusieurs

---

<sup>26</sup> L'expression « modèles reposant sur l'abstraction » (« abstraction models ») réfère aux prototypes.

<sup>27</sup> Il écrit (1990, 65) : « one cannot distinguish between exemplar and abstracted representations in terms of the information they contain. Either type can contain or discard idiosyncratic information. »

propriétés. Par exemple, un prototype de chien pourrait inclure la connaissance que les propriétés *avoir une tache noire sur le bout de la langue, aboyer et avoir une queue de 3,5 cm* (etc.) sont coinstanciées une seule fois. Cela revient à encoder la connaissance qu'il y a un unique individu qui possède ces propriétés. Ainsi ce prototype semble encoder une connaissance à propos d'un individu singulier, comme le ferait un exemplaire.

Pareillement, dans certains modèles des exemplaires, les exemplaires peuvent encoder toutes les connaissances générales qu'un prototype pourrait encoder (voir les exemples dans Barsalou 1990, figures 2.1., 2.2. et 2.3.). Cette idée est bien connue : puisque de nombreux exemplaires sont stockés en MLT, l'esprit peut utiliser les propriétés représentées par tous les exemplaires, afin de déterminer quelles propriétés sont typiques de leur catégorie. Si un exemplaire est stocké en MLT pour chaque individu rencontré, l'ensemble des exemplaires peut encoder toutes les connaissances qu'un prototype abstrait à partir de ces individus pourrait encoder.

J'appelle ces modèles des prototypes et des exemplaires des modèles hybrides, puisqu'ils encodent à la fois des connaissances à propos des propriétés typiques des catégories et des connaissances à propos des individus. *Selon un modèle hybride des prototypes, un prototype encode les connaissances que les exemplaires encodent et selon un modèle hybride des exemplaires, un ensemble d'exemplaires encode les connaissances qu'un prototype encode.*

Ce qui distingue ces modèles, *c'est la manière dont les connaissances sont stockées – et non quelles connaissances sont stockées.* Ainsi, dans le modèle des exemplaires décrit ci-dessus, l'ensemble d'exemplaires encode *implicitement* les connaissances à propos des propriétés typiques de la catégorie qu'il représente. Ces connaissances doivent être extraites de cet ensemble d'exemplaires par un processus cognitif. Pareillement, dans le modèle des prototypes décrit ci-dessus, les connaissances à propos d'individus singuliers sont encodées *implicitement*. Elles doivent aussi être extraites par un processus. Autrement dit, il n'y a pas de représentation d'individus singuliers, mais une représentation de la coinstanciation unique de propriétés à partir de laquelle des connaissances sur des individus singuliers peuvent être extraites.

L'équivalence informationnelle entre certains modèles des prototypes et n'importe quel modèle des exemplaires, et entre certains modèles des exemplaires et n'importe quel modèle des prototypes, implique que le PP et le PE ne peuvent être distingués au moyen des décisions de catégorisation des sujets. Ces décisions révèlent seulement quelles connaissances sont utilisées quand les sujets doivent résoudre une tâche expérimentale. Puisque dans certains modèles, les prototypes encodent toutes les connaissances que les exemplaires encodent (et vice-versa), il est impossible d'inférer à partir des décisions de catégorisation si les sujets utilisent des prototypes ou des exemplaires.

Les temps de réaction des sujets semblent être plus utiles pour distinguer le PP et le PE. Car même quand ils sont informationnellement équivalents, prototypes et exemplaires sont supposés encoder différemment les connaissances. Cependant, l'existence de nombreux modèles des processus cognitifs rend les temps de réaction peu informatifs. Car un même modèle des concepts conduit à différentes prédictions à propos des temps de réaction quand il est couplé à différents modèles des processus.

Barsalou conclut donc que le PP et le PE ne peuvent être distingués expérimentalement.

- Les thèses à propos de la nature des concepts ne peuvent être testées indépendamment de modèles à propos des processus cognitifs (Malt 1989, Barsalou 1990, voir le chapitre 1, section 3.4.). Or les modèles des processus cognitifs doivent inclure des modèles spécifiques de l'extraction des connaissances de la MLT (voir, par exemple, Medin et Edelson 1988 pour un modèle où l'extraction joue un rôle important).

Le problème est alors le suivant : on peut concevoir de très nombreux modèles d'extraction des connaissances. Dans la plupart de ces modèles, les connaissances ne sont pas extraites intégralement ou avec une entière fiabilité. Elles sont extraites partiellement et peuvent être transformées durant l'extraction. Cela implique que les performances des sujets dans les tâches cognitives ne dépendent pas directement des connaissances qui sont stockées dans leurs concepts. Elles dépendent, au contraire, en partie de la nature du processus d'extraction. Barsalou écrit ainsi (1990, 77) :

« To the extent that a processing environment exhibits partial matching, partial retrieval, reconstruction, retrieval failure, and abstraction at retrieval, the output of a model increasingly diverges from information in memory. Because most cognitive theorists assume the presence of such mechanisms in humans, we face much difficulty in identifying representations. We can only observe that information in a form produced by those processes. All category information ever encountered may be stored in memory. But much of it may be unobservable or transformed, given the processing environment.

This state of affairs compounds the difficulty of determining whether people store category knowledge as exemplar memories or as abstractions. »

La position défendue par Barsalou est problématique pour l'ANE. En effet, s'il est impossible de distinguer expérimentalement le PP du PE, les données expérimentales ne peuvent supporter l'existence des exemplaires *et* des prototypes. Car les données expérimentales supposées supporter l'existence des exemplaires (ou des prototypes) pourraient tout aussi bien supporter l'existence des prototypes (ou des exemplaires).

#### *4.2. Réfutation*

Je concède les deux prémisses de Barsalou. Il y a en effet des modèles des prototypes dans lesquels les prototypes encodent les mêmes connaissances que n'importe quel ensemble d'exemplaires et vice-versa. Ces modèles se distinguent seulement par la manière dont les connaissances sont encodées. Il est aussi vrai que l'on ne peut tester les modèles des concepts indépendamment de modèles des processus cognitifs, y compris des modèles de l'extraction des connaissances. Et il est vrai que de nombreuses combinaisons sont possibles. Cependant, je rejette la conclusion de Barsalou : il n'est pas vain de chercher à déterminer si un phénomène psychologique donné est mieux expliqué au moyen d'exemplaires ou de prototypes (ou, pour être complet, de théories).

Considérons d'abord l'argument de l'équivalence informationnelle. La réponse est la suivante. L'équivalence informationnelle entre certains modèles des exemplaires et les modèles des prototypes et entre certains modèles des prototypes et les modèles des exemplaires implique que l'on ne peut distinguer expérimentalement le PP et le PE

seulement si l'on a une raison d'admettre les modèles hybrides. Mais, Barsalou ne nous donne aucune raison de prendre au sérieux ces modèles. Pourquoi donc devrait-on supposer que les entités théoriques qu'ils décrivent peuplent les esprits humains ? Pour quelle raison devrait-on admettre que nos concepts sont adéquatement décrits par un modèle hybride des prototypes ou par un modèle hybride des exemplaires ?

Il faut en effet avoir présent à l'esprit que les modèles des prototypes et les modèles des exemplaires ne sont informationnellement équivalents que si l'on conçoit les prototypes ou les exemplaires d'une manière bien particulière. Les modèles proposés par Barsalou admettent que *les concepts encodent toutes les connaissances possibles à propos des catégories qu'ils dénotent* (Barsalou 1990, 72). Les modèles hybrides des prototypes admettent que les prototypes encodent toutes les coinstanciations de propriétés et les modèles hybrides des exemplaires admettent qu'un exemplaire est formé pour chaque membre rencontré de la catégorie. Sans ces deux hypothèses, il ne serait pas possible d'extraire d'un prototype *toutes* les connaissances qu'un ensemble d'exemplaires pourrait encoder et il ne serait pas possible d'extraire d'un ensemble d'exemplaires *toutes* les connaissances qu'un prototype pourrait encoder. Par exemple, supposons que Paul et Pierre aient rencontré trois A. 2 de ces A,  $A_1$  et  $A_2$ , avaient la propriété F, mais non le troisième  $A_3$ . Paul forme seulement un exemplaire de  $A_1$  et un exemplaire de  $A_2$ . Pierre forme un prototype sur la base de ces trois objets : son prototype encode donc la croyance que deux tiers des A ont la propriété F. Si Paul cherchait à extraire une connaissance générale sur la fréquence de F parmi les A à partir des représentations singulières de  $A_1$  et  $A_2$ , il conclurait au contraire que tous les A ont la propriété F. En conséquence, *si l'on rejette ces deux hypothèses, les modèles des prototypes et les modèles des exemplaires ne sont pas informationnellement équivalents*. Qui plus est, les connaissances qui sont encodées par les exemplaires et par les prototypes diffèrent d'autant plus que les ensembles d'exemplaires incluent peu d'exemplaires et que les prototypes encodent peu de connaissances sur les coinstanciations de propriétés.

Pourquoi rejeter les hypothèses sur lesquelles reposent les modèles hybrides de Barsalou ? Tout d'abord, ces hypothèses sont entièrement implausibles.<sup>28</sup> Il est simplement *faux* que nous formons une représentation singulière pour tous les individus

---

<sup>28</sup> Notons que Barsalou n'admet pas ces modèles. Son argument est de nature sceptique.

que nous rencontrons. Il est aussi faux que nous encodons toutes les coinstanciations de propriétés.

Finalement, la seule raison d'admettre ces modèles hybrides semblent dépendre de l'hypothèse EN. Puisque les concepts sont supposés être homogènes et expliquer *toutes* les propriétés des processus cognitifs, seuls les modèles dans lesquels les concepts encodent à la fois des connaissances sur des propriétés typiques et des connaissances à propos de membres particuliers semblent adéquats.<sup>29</sup> Ces modèles semblent les seuls à pouvoir expliquer l'ensemble des propriétés qui caractérisent nos compétences cognitives supérieures. Par exemple, avant de présenter sa propre théorie des concepts, Prinz écrit (Prinz 2002, 103, voir aussi Barsalou 1990, 84, Smith 1995):

« The theory I propose appropriates elements from almost all of the theories that I reviewed. »

Supposons un instant que l'hypothèse EN soit fautive. Différentes propriétés de nos compétences cognitives peuvent alors être expliquées par différents types de représentations. Par exemple, certaines propriétés de notre compétence inductive peuvent être expliquées par des processus dans lesquels figurent *des prototypes*, tandis que d'autres peuvent être expliquées par des processus dans lesquels *des théories* figurent – ou *des exemplaires*. Bref, nul n'est alors besoin d'admettre l'existence de représentations qui encoderaient implicitement ou explicitement toutes les connaissances à propos des propriétés typiques, des membres individuels etc.

---

<sup>29</sup> La théorie des concepts proposée par illustre cette dialectique. Selon Prinz, les concepts doivent expliquer toutes les propriétés de nos processus conceptuels. Il admet ainsi qu'un unique type de représentations mentales est impliqué dans nos processus conceptuels et qu'il explique leurs propriétés. En d'autres mots, il accepte l'hypothèse EN. Pour expliquer toutes les propriétés des processus conceptuels, il est conduit à admettre une théorie hybride des concepts qui additionnent les paradigmes traditionnels des concepts (voir le chapitre 3, section 3.). Il écrit ainsi (2002, 164) : « The theory of concepts outlined in this chapter *integrates the best features of other theories*. It borrows the view that concepts are perceptually based from imagism. It borrows a theory of informational content from informational atomism. It borrows the idea of instance information and essentialist beliefs from exemplar theory and the theory theory. It borrows idealized summary representations from prototype theory. *Proxytype theory is a hybrid*, but its many facts can be unified under a single overarching idea : concepts are mechanisms that allow us to enter into perceptually mediated, intentionality conferring, causal relations with categories in the world. » Je souligne.

En conséquence, quand l'on rejette l'hypothèse EN, il ne semble rester aucune raison d'admettre les modèles hybrides de Barsalou. Les autres modèles des prototypes et des exemplaires ne sont pas informationnellement équivalents. En conséquence, nous n'avons aucune raison de croire qu'ils ne peuvent être expérimentalement distingués.

Considérons maintenant le second argument. Il est vrai que l'on ne peut tester un modèle des concepts indépendamment des modèles des processus cognitifs. Il est aussi vrai que les données sur les performances de catégorisation des sujets, y compris les temps de réaction, ne sont peut-être pas en mesure de distinguer tous les modèles possibles des concepts.

La nature des mesures dépendantes qui sont habituellement utilisées dans les tâches de catégorisation aggrave ce problème. Les psychologues mesurent habituellement certains aspects des *résultats* du processus de catégorisation. En conséquence, les mesures dépendantes sont le reflet des différentes étapes du processus de catégorisation (extraction de l'information, calcul de ressemblance etc.). Ce type de mesure offre donc au mieux un témoignage indirect sur la nature des représentations utilisées dans le processus de catégorisation. C'est l'élément central de la seconde objection de Barsalou.

J'offre trois réponses. Tout d'abord, il y a de nombreuses mesures des performances de catégorisation. Certaines sont peu utilisées dans la psychologie contemporaine de la catégorisation. Malt a proposé avec raison d'utiliser un spectre plus large de mesures dépendantes, y compris des effets de facilitation (« priming ») et des comptes-rendus conscients (Malt 1989, voir le chapitre 5, section 4.4.). Ces mesures peuvent être un reflet plus direct de la nature des représentations utilisées dans le processus de catégorisation.<sup>30</sup> En outre, les théories des concepts ne sont pas limitées aux performances de catégorisation (comme le remarque correctement Barsalou dans sa conclusion). La plupart des processus cognitifs de haut niveau peuvent fournir des données empiriques qui peuvent être utilisées pour évaluer différentes théories des

---

<sup>30</sup> Malt écrit (1989, 540): « A more fruitful approach at this point may be to move away from measures reflecting the end product of the categorization and instead to seek measures that reflect ongoing processes. Information about any aspect of the events leading up to the classification decision can help constrain theories of categorization. If evidence can be found for use of a particular process of retrieval of a particular kind of information while a categorization decision is being made, some classes of models will be supported and others ruled out. »

concepts (voir le principe de promiscuité, chapitre 1, section 2.2.). Finalement, des considérations théoriques sont aussi pertinentes : par exemple, des considérations de simplicité peuvent être prises en considération. Si nous prenons en compte toutes les données empiriques et tous les arguments théoriques, certains modèles sont sans doute plus plausibles que d'autres.

Je conclus donc que les arguments de Barsalou n'établissent pas sa conclusion. *Les modèles crédibles des prototypes et des exemplaires sont informationnellement différents et peuvent être expérimentalement distingués.*

#### 4.3. Conclusion

Barsalou a soutenu que le PP et le PE ne pouvaient être distingués expérimentalement. Cependant, ses arguments ne supportent pas cette thèse provocatrice. Ils se réduisent aux deux points suivants. Tout d'abord, il y a des modèles des prototypes que l'on ne peut distinguer expérimentalement des modèles des exemplaires (et vice-versa), parce qu'ils sont informationnellement équivalents. Cependant, nous avons vu que ces modèles reposaient sur des hypothèses implausibles. Deuxièmement, les mesures des performances de catégorisation sont le reflet de plusieurs variables en sus de la nature des représentations utilisées. Cependant, nous avons vu que différentes mesures à propos de plusieurs compétences cognitives supérieures et des considérations théoriques pouvaient être utilisées pour distinguer les différents modèles des concepts. En conséquence, les théories spécifiques des concepts inspirées par les trois paradigmes étudiés dans ce travail ne sont pas expérimentalement équivalentes.

### 5. Conclusion

La prémisse 3 de l'ANE est cruciale : pour que l'argument de la nécessité explicative soit convaincant, il faut montrer que les concepts se divisent en types qui ont peu en commun. Ma stratégie est la suivante : je montre que les paradigmes des concepts ont postulé des types de représentations qui ont peu en commun. Si je réussis à montrer qu'il faut postuler l'existence de tous ces types de représentations mentales, je peux alors conclure que les concepts ne forment pas une espèce naturelle.

Les principaux paradigmes des concepts ont postulé trois entités théoriques, les prototypes, les exemplaires et les théories (voir le chapitre 3). Il est assez clair qu'ils diffèrent. Mais cela est insuffisant pour établir que les concepts ne forment pas une espèce naturelle. Pour supporter le rejet de l'hypothèse EN, il faut montrer que ces trois types de représentations ont peu en commun. Est-ce le cas ?

La réponse dépend de notre théorie à propos de la nature des représentations : quelles propriétés sont importantes pour caractériser un type de représentations ? Or, les représentations sont habituellement caractérisées par trois types de propriétés, leurs propriétés sémantiques, leurs propriétés fonctionnelles et les propriétés de leurs véhicules. J'ai montré que les trois grands paradigmes des concepts supposaient que les prototypes, les théories et les exemplaires avaient des propriétés sémantiques différentes. Les prototypes expriment des propositions statistiques, les exemplaires, des propositions singulières et les théories, des propositions théoriques. Prototypes, exemplaires et théories ont aussi des propriétés fonctionnelles différentes. Les prototypes et les exemplaires sont supposés figurer dans des processus qui reposent sur des calculs de ressemblance. Mais le PP et le PE caractérisent différemment la nature de la ressemblance. Les théories sont supposées figurer dans des processus fondés sur des règles. Finalement, nous connaissons trop peu sur les véhicules des concepts pour que leurs propriétés soient prises en considération pour évaluer la ressemblance entre prototypes, exemplaires et théories. Je conclus donc que les prototypes, les exemplaires et les théories sont trois types de représentations mentales qui ont peu en commun. J'ai donc établi l'énoncé conditionnel suivant : si l'on peut montrer, sur la base de données empiriques (voir chapitres 5-7) que les concepts se divisent en prototypes, exemplaires et théories, alors la classe des concepts se divise en types de représentations qui ont peu en commun.

L'on pourrait concéder cet énoncé conditionnel, tout en niant que l'on puisse jamais distinguer expérimentalement ces paradigmes. Barsalou a offert deux arguments qui vont dans ce sens. J'ai cependant montré que ce pessimisme n'était nullement justifié. Le PP, le PE et le PT posent donc des entités qui ont peu en commun et que l'on peut chercher à distinguer expérimentalement.

## Chapitre 5

### L'hétérogénéité de la catégorisation

---

Les chapitres cinq à huit ont pour objet de justifier la prémisse 2 de l'argument de la nécessité explicative (ANE) : il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs. La prémisse 2 soutient en bref qu'il y a de bonnes raisons d'admettre au sein de notre ontologie des prototypes, des exemplaires et des théories (ainsi, peut-être, que d'autres types de représentations mentales). Je montre dans ces trois chapitres que l'hypothèse selon laquelle des exemplaires, des prototypes et des théories sont utilisés par les compétences cognitives supérieures humaines est la meilleure explication de nombreuses propriétés de ces compétences.

Le chapitre 5 porte sur une capacité cognitive très importante, *la compétence de catégorisation*. Il est difficile de surestimer le rôle qu'a joué et que joue encore cette compétence dans l'étude empirique des concepts. Car la catégorisation a été pendant plus de trente ans, et, dans une large mesure, est toujours le principal champ de bataille entre les approches concurrentes des concepts. La plupart des théories des concepts sont avant tout des théories de la catégorisation : elles ont été constamment testées au moyen de données obtenues dans des tâches de catégorisation et leur validité empirique a souvent été mesurée à leur capacité à rendre compte des performances de catégorisation des êtres humains.<sup>1</sup>

---

<sup>1</sup> Certains psychologues, particulièrement les avocats du PT, se sont plaints de l'importance accordée à la catégorisation. Ainsi, en psychologie développementale, Gelman et Coley écrivent (1991, 162, voir aussi Gelman et Medin 1993) : « past work typically assumed that category knowledge can best be tapped by asking children to sort or classify objects. Tasks used previously most often required children to construct a classification (e.g., « put together the things that go together ») – thus emphasizing similarity – rather than to reason about or draw inferences from a category – thereby emphasizing the non-obvious. When we ask children simply which objects belong together, we are neglecting the deeper questions of whether the grouped objects form a motivated category. It seems clear that concepts comprise more knowledge than knowledge of which instances belong in a category (...). Indeed (...) inductions about completely new instances and properties reveal important aspects of category function. » Dans la recherche contemporaine sur les concepts, la catégorisation joue encore un rôle très important. Ainsi Prinz inclut-il la catégorisation parmi les explananda des théories des concepts, alors qu'il ne mentionne pas les autres compétences

Comme les théories des concepts, associées à des théories plus ou moins précises des processus cognitifs, doivent expliquer les propriétés de nos performances de catégorisation (voir le principe de l'efficacité causale, chapitre 1, section 2.3., et la méthodologie habituelle, chapitre 1, section 3.4.), les psychologues ont souvent cherché à découvrir de manière expérimentale des propriétés de la catégorisation qui pouvaient être expliquées par leurs théories préférées, mais non par les théories concurrentes – contraignant celles-ci à embrasser des hypothèses *ad hoc*.<sup>2</sup>

Or, on doit reconnaître qu'aucune théorie des concepts n'explique de nos jours *toutes* les découvertes à propos de la catégorisation. Chaque théorie explique remarquablement un sous-ensemble de ces données, mais est souvent incapable d'expliquer les autres. Cette situation a conduit certains psychologues à douter de l'adéquation des principaux paradigmes des concepts (Smith et Minda 2000, Murphy 2002). Cependant, comme nous allons le voir, ce n'est pas la seule ni même la plus plausible conclusion que suggère cette situation.

Le but de ce chapitre est en effet de mettre cette situation sous un nouveau jour. Cette situation ne montre pas que les principaux paradigmes actuels des concepts, le paradigme postulant des prototypes (PP), le paradigme postulant des exemplaires (PE) et le paradigme postulant des théories (PT), sont inadéquats et doivent être à terme remplacés par un nouveau paradigme des concepts ; elle ne montre pas non plus que les données expérimentales sur la catégorisation ne sont pas adéquates pour distinguer les différentes théories des concepts (Barsalou 1990, Murphy 2002, voir le chapitre précédent, section 3.). Je propose que l'incapacité des différentes théories des concepts et de la catégorisation à expliquer l'ensemble des données empiriques résulte de la fausseté de l'hypothèse EN. Plus précisément, je soutiens dans ce chapitre la thèse suivante :

### **L'hétérogénéité de la catégorisation**

(HC) La compétence de catégorisation des êtres humains est réalisée par plusieurs processus cognitifs et chaque processus cognitif implique un type

---

cognitives supérieures (Prinz 2002). Il caractérise aussi nos concepts comme des mécanismes de détection : cela revient à lier essentiellement les concepts à la compétence de catégorisation.

<sup>2</sup> Le PC fut, par exemple, rejeté parce qu'il était incapable d'expliquer les phénomènes de typicalité.

différent de représentations, y compris des prototypes, des exemplaires et des théories.<sup>3</sup>

Si la thèse HC est vraie, les théories des concepts qui ne posent qu'un seul type de concepts, comme la plupart des théories contemporaines, ne peuvent, par principe, expliquer *toutes* les données empiriques à propos de la catégorisation – ce qui explique la situation décrite ci-dessus.

Dans une première section, j'explique la notion de catégorisation : que veulent dire les psychologues par « catégorisation » ? Dans une deuxième section, je présente trois modèles de notre compétence de catégorisation, inspirés par les trois principaux paradigmes des concepts. Dans une troisième section, je présente les données empiriques qui montrent que certaines performances de catégorisation impliquent des concepts théoriques. Dans la quatrième section, je montre qu'il faut aussi postuler l'existence de prototypes et d'exemplaires pour expliquer nos performances de catégorisation. Dans la dernière section, je compare HC avec d'autres explications possibles des données empiriques sur la catégorisation.

## 1. Qu'est-ce que la catégorisation ?

Dans l'introduction de ce chapitre, je n'ai pas caractérisé la compétence de catégorisation. Je m'efforce de le faire dans cette section.<sup>4</sup> Je montre aussi qu'il est plausible que notre compétence de catégorisation soit réalisée par plusieurs processus cognitifs qui reposent sur différents types de représentations mentales.<sup>5</sup>

---

<sup>3</sup> Cette idée n'est pas entièrement nouvelle. Elle a été en partie défendue par les partisans des théories duales des concepts (par exemple, Osherson et Smith 1981, voir le chapitre 2, section 3.2., et ci-dessous section 5.). Elle a aussi été défendue par quelques psychologues dans les années 1980, par exemple Homa et al. 1981, Medin et Edelson 1988, Malt 1989, Ross et al. 1990. De nos jours, de plus en plus de psychologues admettent HC sous une forme ou sous une autre (par exemple, Ashby et al. 1998, Hahn et Chater 1998, Smith et Minda 1998, Smith et al. 1998, Smith et Jonides 1999). Cependant, ce n'est pas encore l'approche dominante.

<sup>4</sup> Je propose donc une hypothèse interprétative de la notion de catégorisation dans la littérature psychologique.

<sup>5</sup> Sur la distinction compétence / processus, voir le chapitre 1, section 1.2.1., et le chapitre 4, section 2.5.

### 1.1. La compétence de catégorisation

Que veulent dire les psychologues par « catégorisation » ? Il est étrange de constater qu'habituellement, ils ne décrivent pas précisément la compétence qui est désignée par la notion de catégorisation. Essayons de combler ce vide. Selon la méthodologie classique en sciences cognitives (Marr 1982, voir la discussion de Peacocke 1986, Clark 1993 et Machery 2004), décrire une compétence cognitive, c'est spécifier une fonction qui transforme un ensemble d'inputs en un ensemble d'outputs.<sup>6</sup> La question est donc : *quels sont les inputs et les outputs de notre compétence de catégorisation ?*

Le principal type d'outputs de notre compétence de catégorisation consiste en des *jugements d'appartenance* : un individu est jugé appartenir à une classe. Ces jugements ont la forme suivante :

$$a \in \{x: x \text{ est un } F\}$$

Cependant, les psychologues ne limitent pas les outputs de notre compétence de catégorisation à ce type de jugements (par exemple, Rosch et Mervis 1975, Hampton 1979, Murphy et Medin 1985...). En effet, notre compétence de catégorisation produit aussi selon eux des *jugements d'inclusion* : une sous-classe est jugée être incluse dans une autre classe. Ces jugements ont la forme suivante :

$$\forall x (x \in \{y: y \text{ est un } F\} \rightarrow x \in \{y: y \text{ est un } G\})$$

Les psychologues ne distinguent pas toujours explicitement ces deux types de jugements (Hampton 2001 est une exception), bien que leurs formes soient différentes. Ils semblent admettre que notre compétence de catégorisation produit ces deux types de jugements. Il se peut que la plupart des psychologues ignorent simplement la différence

---

<sup>6</sup> Cette fonction peut être spécifiée plus ou moins précisément. Elle pourrait être spécifiée en détail si l'on connaissait tous les n-tuplés (argument, valeur) qui caractérisent la fonction. Elle est habituellement spécifiée beaucoup moins précisément : elle est caractérisée comme transformant un type d'inputs en un type d'outputs. Par exemple, notre compétence visuelle est décrite comme suit : elle transforme des patterns d'excitation des cônes et des bâtons de notre rétine en représentations tridimensionnelles de l'environnement distal. Dans ce chapitre, la compétence de catégorisation est semblablement décrite.

entre ces deux types de jugements. En conséquence, ils pourraient simplement ne pas voir qu'un jugement d'appartenance et un jugement d'inclusion pourraient être cognitivement différents. On peut cependant proposer une interprétation plus charitable – et plus probable. Les psychologues comprennent la différence entre ces deux types de jugements. Cependant, pour des raisons théoriques et empiriques, ils soutiennent qu'ils sont cognitivement semblables. Cette idée est plausible. Car, d'un point de vue empirique, les jugements d'inclusion et les jugements d'appartenance ont des propriétés semblables. Par exemple, on trouve des effets de typicalité dans les deux cas. Médor est un chien typique et les moineaux sont des oiseaux typiques. Bien plus, comme nous allons le voir, les modèles des processus qui produisent des jugements d'appartenance sont à même d'expliquer aussi les jugements d'inclusion. Il se peut donc que ces deux types de jugements résultent d'un unique processus cognitif : prendre une décision d'appartenance et prendre une décision d'inclusion reposeraient sur le même processus psychologique. Ainsi les jugements d'inclusion et d'appartenance constitueraient-ils les outputs d'une unique compétence. Dans ce qui suit, j'admets cette seconde interprétation.<sup>7</sup>

Pour résumer, dans la littérature psychologique, « catégorisation » réfère à une compétence qui produit des jugements d'appartenance et d'inclusion. Prendre une décision d'appartenance et prendre une décision d'inclusion sont supposés être le même processus psychologique.

Quels sont les inputs de notre compétence de catégorisation ? De nouveau, les psychologues ne sont pas très clairs sur cette question. Cependant, les expériences qui sont utilisées pour tester les théories de la catégorisation peuvent nous éclairer. En effet, les psychologues utilisent souvent la procédure expérimentale suivante : dans une phase d'apprentissage, on présente aux sujets des objets qui appartiennent à une ou plusieurs catégories ; on présente de nouveaux objets, les cibles dans la phase de test.<sup>8</sup> Sur la base

---

<sup>7</sup> On pourrait faire l'objection suivante. Les jugements d'appartenance et d'inclusion correspondent en fait à deux compétences différentes. Simplement, elles sont réalisées par le même processus. L'individuation des compétences est un réel problème. Cependant, j'admets ici que deux compétences ne peuvent être réalisées par un unique processus.

<sup>8</sup> La phase de test utilise parfois les objets présentés dans la phase d'apprentissage (par exemple, Rosch et Mervis 1975).

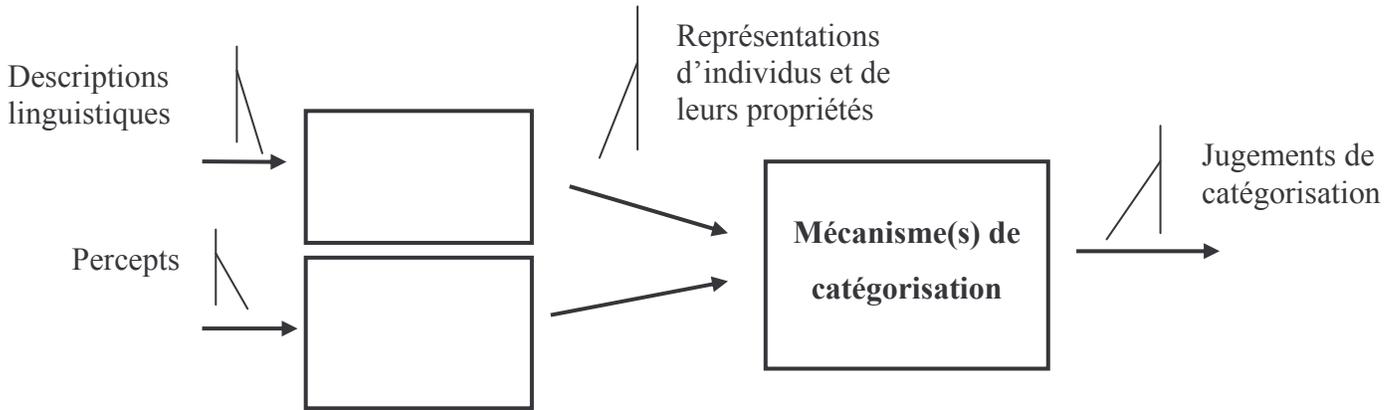
des objets présentés dans la phase d'apprentissage, les sujets sont supposés former un concept de la catégorie ou des catégories. Dans la phase de test, ils sont supposés utiliser ce concept pour catégoriser les cibles.

On peut inférer la nature supposée des inputs de la compétence de catégorisation à partir de la nature des objets utilisés dans les phases d'entraînement et de test. Dans ces expériences, différents types d'objets sont présentés. Certains objets sont *des entités artificielles*, par exemple des patterns de points (Posner et Keele 1968, 1970, Knowlton et Squire 1993), des suites de nombres (par exemple, Rosch et Mervis 1975) ou des dessins (Wisniewski et Medin 1994). D'autres objets sont *des entités non artificielles*. Les entités non artificielles, le plus souvent des objets de taille moyenne, peuvent être présentées *perceptuellement*, habituellement au moyen de dessins plus ou moins réalistes ou de photos (voir, par exemple, Wisniewski et Medin 1994 et les expériences décrites dans Sloman et al. 2001). Les entités non artificielles peuvent aussi être présentées *verbalement* : les objets peuvent être décrits aux sujets (voir les expériences présentées dans Ahn et Dennis 2001 et dans Hampton 2001). Ces descriptions peuvent être plus ou moins longues : elles décrivent parfois deux ou trois propriétés de l'objet (voir, par exemple, les descriptions utilisées dans les expériences de Ahn et Dennis 2001), mais sont aussi parfois plus développées. Les descriptions peuvent mentionner uniquement des propriétés perceptuelles ou d'autres propriétés. Finalement, dans les expériences qui cherchent à étudier les jugements d'inclusion, on présente souvent aux sujets des mots, par exemple « moineau » et les sujets doivent prendre une décision de catégorisation, par exemple répondre à la question « est-ce qu'un moineau est un oiseau ? » (Rips et al. 1973).

Les cibles et les objets utilisés lors de l'entraînement n'appartiennent donc pas à une modalité spécifique particulière, même si de nombreuses expériences utilisent des stimuli visuels. En conséquence, la compétence de catégorisation n'est pas liée à une modalité perceptuelle spécifique. En outre, les cibles et les objets utilisés lors de l'entraînement n'ont pas à être perceptuels, puisque de nombreuses expériences utilisent des inputs linguistiques. Ainsi la compétence de catégorisation n'est-elle pas essentiellement liée à la perception.

On peut donc faire la proposition suivante. Sur la base des perceptions ou des inputs linguistiques, l'esprit produit des représentations d'objets ou de classes. Ces

représentations attribuent des propriétés à des objets ou à des classes. Ces représentations sont les inputs du ou des systèmes de catégorisation. Il est plausible que différents mécanismes contribuent à produire ces représentations (voir la discussion dans Machery 2004). Par exemple, les descriptions linguistiques et les percepts déclenchent probablement différents mécanismes. Le schéma suivant résume cette idée.



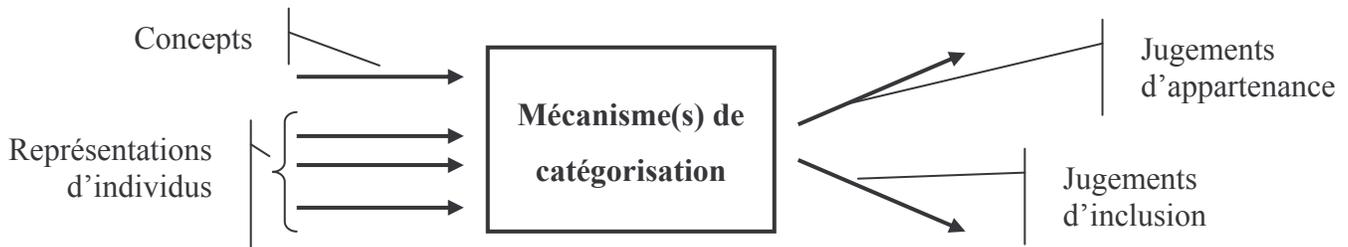
**Schéma 5.1.**

Nous pouvons donc caractériser les inputs de la compétence de catégorisation comme suit.

- Tout d'abord, ils incluent les représentations qui attribuent des propriétés spécifiques à des objets particuliers. Ces représentations peuvent être produites à partir d'informations perceptuelles ou linguistiques.
- Ils incluent aussi des représentations de classes, c'est-à-dire des concepts. Quand on demande aux sujets de former des jugements d'inclusion, les noms utilisés dans les expériences activent des représentations des classes dénotées. Ces représentations servent d'inputs aux processus de catégorisation.

En résumé, la compétence de catégorisation semble être la capacité de produire des jugements d'appartenance à partir de représentations mentales qui attribuent des propriétés à des individus et des jugements d'inclusion à partir de représentations mentales qui décrivent des classes d'objets. Les représentations mentales qui servent

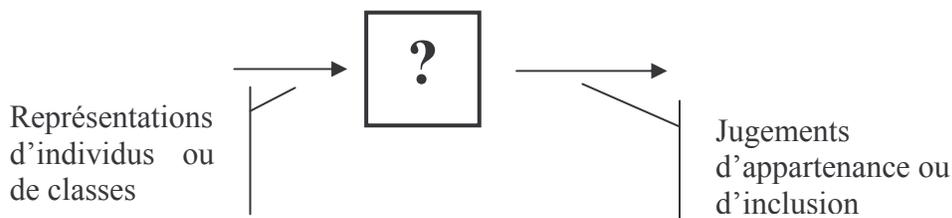
d'inputs peuvent être produites à partir de perceptions, de descriptions linguistiques ou peuvent être extraites de la MLT. Le schéma suivant résume cette idée.



**Schéma 5.2.**

### 1.2. Les processus de catégorisation

Etant donné cette caractérisation de la compétence de catégorisation, on peut se poser la question suivante : *quels sont les mécanismes qui réalisent cette compétence ?* Considérons le schéma suivant :



**Schéma 5.3.**

Le but est de rendre compte du point d'interrogation. Les théories de la catégorisation sont supposées s'atteler à cette tâche. Certaines théories offrent des modèles détaillés des mécanismes computationnels supposés réaliser la compétence de catégorisation ; d'autres offrent des caractérisations plus qualitatives et moins précises.

Remarquons que la compétence de catégorisation est une compétence *psychologique* et que l'explication de cette compétence doit être psychologique. Les psychologues cherchent à décrire les processus cognitifs qui produisent des jugements de catégorisation. Leur tâche n'est pas sémantique : ils n'essaient pas d'expliquer pourquoi

un concept réfère à telle classe etc. Ils cherchent à expliquer pourquoi et comment nous prenons les décisions de catégorisation que nous prenons.<sup>9</sup>

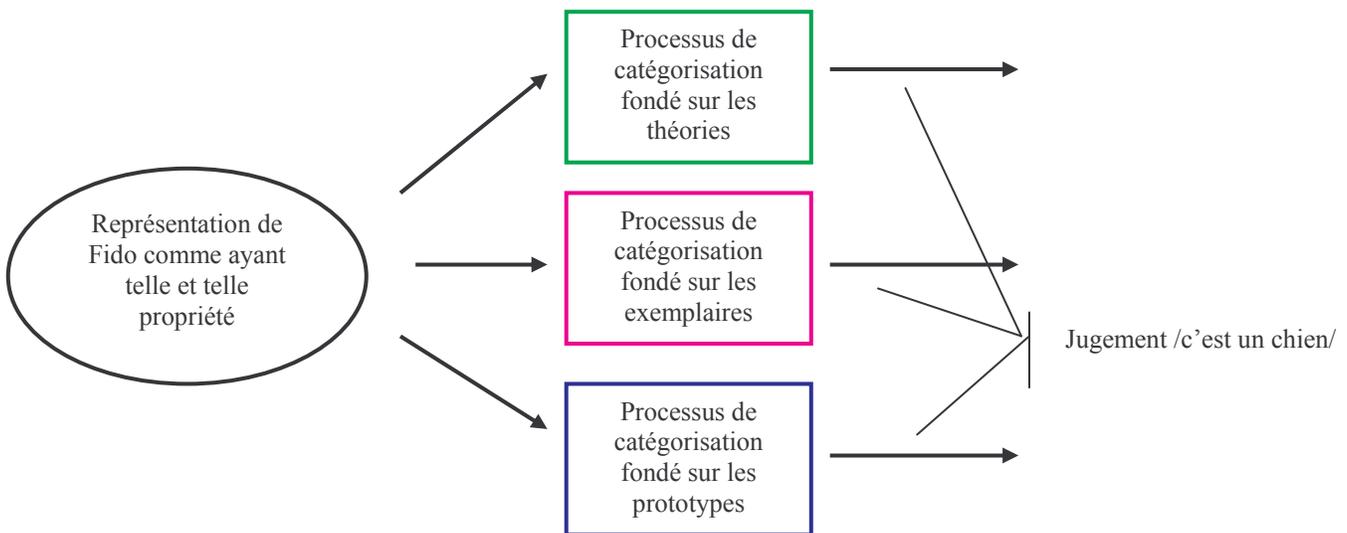
### 1.3. L'hétérogénéité de la catégorisation (HC)

Dans ce chapitre, je cherche à montrer que la littérature sur la catégorisation vient à l'appui de l'idée selon laquelle *notre compétence de catégorisation est réalisée par plusieurs processus cognitifs qui impliquent des types différents de représentations.*

Mon hypothèse est double :

- Il existe plusieurs processus qui peuvent produire un jugement de catégorisation donné à partir d'un input donné. Par exemple, supposons que Pierre perçoit un chien. Pierre produit le jugement /c'est un chien/.<sup>10</sup> L'idée est qu'il existe plusieurs processus qui peuvent produire ce jugement à partir de la perception de Pierre.
- Il existe au moins trois processus de catégorisation différents, chacun impliquant un type différent de représentations, à savoir des exemplaires, des prototypes et des théories.<sup>11</sup>

Le schéma suivant exprime cette idée.



**Schéma 5.4.**

<sup>9</sup> De nombreux philosophes négligent cette distinction quand ils discutent les théories de la catégorisation (Rey 1983, 1985, Fodor 1994, 1998).

<sup>10</sup> J'utilise // pour nommer les jugements.

<sup>11</sup> Je ne soutiens pas que la compétence de catégorisation soit réalisée seulement par ces trois mécanismes. Mais je soutiens qu'elle est réalisée au moins par ces trois mécanismes.

C'est une version de *la théorie pluraliste des processus cognitifs* que j'ai présentée dans le chapitre 4 (section 2.3.). Nos performances de catégorisation peuvent être les outputs de plusieurs processus cognitifs différents qui sont appliqués aux mêmes inputs.

Cette hypothèse a été soutenue par quelques psychologues. En particulier, Allen et Brooks 1991, Smith, Langston et Nisbett 1992, Smith et Sloman 1994 et Smith et al. 1998 ont défendu des thèses plus ou moins semblables à HC.<sup>12</sup> Je crois que les données empiriques viennent fortement à l'appui de HC.

#### *1.4. Un argument pour l'hétérogénéité de la catégorisation*

HC est une hypothèse très plausible. Car nos performances de catégorisation sont très *variées*. Cette variété suggère que notre compétence de catégorisation est réalisée par plusieurs mécanismes.<sup>13</sup>

L'argument est le suivant. Nos performances de catégorisation sont très diverses. Comme nous l'avons vu, les psychologues admettent que nous catégorisons *une grande variété de stimuli* – ce qui est bien le cas ! En outre, nous catégorisons dans *des circonstances très variées*. *La durée* de nos performances de catégorisation est aussi très diverse : elle varie de quelques centaines de millisecondes, par exemple quand nous déterminons le sexe d'un individu, à des jours ou des semaines, par exemple quand un juge conclut qu'un accusé est coupable (sur la catégorisation ultrarapide, voir Fabre-Thorpe et al. 2000, Thorpe et al. 2001). Certains processus sont *guidés par les stimuli* (« stimuli-driven »), tandis que d'autres ne le sont pas. Dans le premier cas, la cible est présentée perceptuellement et la catégorisation est déterminée presque exclusivement par ces propriétés perceptuelles. Ce type de catégorisation est souvent appelé catégorisation

---

<sup>12</sup> Voir en particulier Smith et al. 1998. Ils écrivent (168) : « In this article we argue that people can apply multiple procedures of categorization to the same objects, perhaps even simultaneously, and that the procedures are qualitatively different from each other. »

<sup>13</sup> Etant donné la diversité de nos performances de catégorisation, on pourrait même douter que notre compétence de catégorisation soit une compétence psychologique unique : il se peut que les psychologues aient à tort assimilé différentes compétences. Je vais cependant admettre ici qu'il est pertinent de traiter la catégorisation comme une compétence unique. Cette remarque soulève une question délicate : comment décide-t-on si une compétence est réellement une ou si c'est en fait un agrégat de plusieurs compétences (voir les remarques sur ce sujet dans Machery 2004 et la note 7 ci-dessus) ?

perceptuelle. De toute évidence, l'exemple du plongeur ivre de Murphy et Medin (voir le chapitre 4, section 2.4.) est très différent. *La certitude subjective* des jugements de catégorisation varie aussi. Certaines catégorisations peuvent dépendre *de structures innées*, tandis que d'autres peuvent dépendre *de nos capacités d'apprentissage*. Finalement, nos performances de catégorisation ont des *fonctions différentes*. Souvent, nous catégorisons en vue d'accroître notre connaissance, tandis que parfois nous catégorisons dans le cours de l'action.

Une telle diversité, en particulier la variété des propriétés temporelles et la différence entre les processus guidés par les stimuli et les autres processus, suggère que la catégorisation dépend de plusieurs processus. Cela vient à l'appui du premier élément de HC (voir l'introduction et la section 1.3. ci-dessus). Car, s'il y a plusieurs processus cognitifs de nature différente, on peut s'attendre à une diversité des performances de catégorisation.

Mais qu'en est-il du second élément de HC, l'idée que parmi ces processus on trouve un processus qui fait appel aux exemplaires, un autre qui fait appel aux prototypes et un dernier qui fait appel aux théories ? Il est certainement possible d'expliquer ainsi plusieurs différences entre nos catégorisations. C'est sans doute le cas des différences temporelles : les jugements qui font appel à des théories sont sans doute plus lents que les jugements qui font appel aux prototypes et aux exemplaires. On peut sans doute aussi expliquer le fait que certains jugements sont guidés par les stimuli, mais non d'autres. Par exemple, la catégorisation fondée sur des théories dépend peut-être d'un contrôle métacognitif.

On pourrait aussi se demander si admettre trois processus est suffisant pour expliquer cette diversité. Cependant, je ne soutiens pas que nous ayons seulement trois processus, mais au moins trois processus (voir Berretty et al. 1999 pour un autre processus possible, la catégorisation par élimination). Finalement, certaines différences entre nos performances de catégorisation ne résultent sans doute pas de l'existence de plusieurs processus, mais des conditions d'utilisation des processus.

### 1.5. Conclusion

Dans cette section, j'ai caractérisé la compétence de catégorisation postulée par les psychologues. Les psychologues se dispensent souvent d'expliquer clairement ce qu'est la catégorisation. Je propose l'explication suivante : c'est la compétence qui produit des jugements d'appartenance à partir de représentations qui attribuent des propriétés à des individus et des jugements d'inclusion à partir de représentations de classes. J'ai admis que ces deux types de jugements dépendent des mêmes mécanismes. J'ai aussi présenté l'hypothèse de l'hétérogénéité de la catégorisation : plusieurs processus qui impliquent différents types de représentations (prototypes, exemplaires et théories) réalisent la compétence de catégorisation. J'ai finalement souligné la plausibilité de HC sur la base de la diversité des performances de catégorisation.

## **2. Les théories de la catégorisation**

Dans le chapitre 4, j'ai présenté rapidement certains aspects des théories de la catégorisation associées aux différents paradigmes des concepts, le PP, le PE et le PT (sections 2.3. et 2.4.). Dans cette section, je présente ces théories plus en détail. Il est important d'avoir présent à l'esprit que les modèles présentés ci-dessous décrivent en fait les principales propriétés de *classes* de modèles. Aucun d'entre eux n'est un modèle complet de la catégorisation.

### *2.1. Les modèles de la catégorisation fondés sur les prototypes et sur les exemplaires*

Je présente ci-dessous deux modèles des décisions d'appartenance. Le premier repose sur des prototypes et le second sur des exemplaires.<sup>14</sup> Ces modèles peuvent être aisément modifiés pour expliquer les décisions d'inclusion.

---

<sup>14</sup> Smith et al. (1998) présentent un modèle un peu différent de la catégorisation fondée sur des exemplaires.

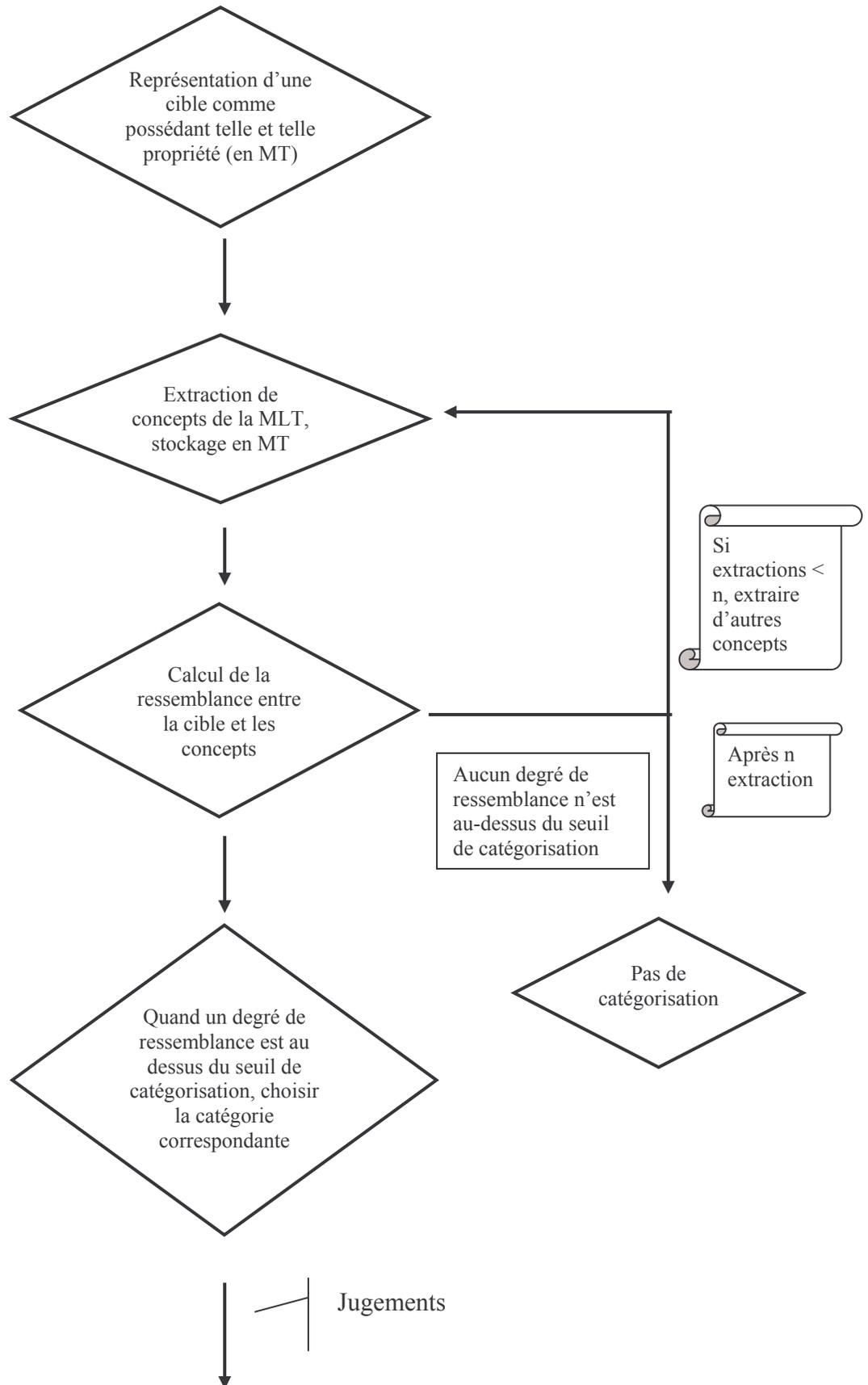


Schéma 5.5.

La première étape est *l'étape d'attribution* (Machery 2004). Il s'agit de produire une représentation de la cible comme possédant plusieurs propriétés : par exemple, ceci aboie, a un pelage, bouge etc.<sup>15</sup> Cette représentation peut être dérivée de notre perception (par exemple, nous pouvons avoir vu une entité qui a ces propriétés) ou de notre compréhension linguistique (nous pouvons avoir lu quelque chose à propos de cette entité). Cette représentation sert d'input aux mécanismes de catégorisation.

La seconde étape est *l'étape de l'extraction*. Elle consiste à extraire des concepts de la MLT et à les stocker en MT (Medin et Edelson 1988, Barsalou 1990). Les représentations qui sont stockées en MT peuvent être dépendantes du contexte (« context-specific ») et peuvent varier d'une extraction à l'autre (voir, par exemple, Barsalou 1999, Prinz 2002, et le chapitre 3, section 3.). Les processus d'extraction ne sont pas spécifiés dans le schéma proposé (pour certaines suggestions, voir Barsalou 1990).<sup>16</sup>

Les modèles fondés sur les exemplaires supposent qu'un ou plusieurs ensembles d'exemplaires (éventuellement dépendants du contexte) sont extraits de la MLT.<sup>17</sup> Les modèles fondés sur des prototypes supposent qu'un ou plusieurs prototypes (éventuellement dépendants du contexte) sont extraits de la MLT. Il est plausible que différents processus d'extraction doivent être imaginés pour ces deux types de représentations. Dans le modèle présenté ci-dessus, plusieurs prototypes/ensembles d'exemplaires sont extraits de la MLT.

La troisième étape est souvent appelée *l'étape de comparaison* (« matching step »). Le système compare les propriétés représentées par la représentation de la cible et les propriétés représentées par le(s) concept(s). Les modèles fondés sur les exemplaires et

---

<sup>15</sup> Les propriétés n'ont pas à être exprimables dans une langue naturelle.

<sup>16</sup> On peut toutefois se demander pourquoi tel concept est extrait de la MLT, et non tel autre. Cela varie bien sûr d'un modèle d'extraction des représentations mentales à l'autre. Dans les modèles connexionnistes, *toutes* les représentations mentales sont extraites à chaque catégorisation. On pourrait aussi proposer que la possession par la cible de certaines propriétés détermine la classe des concepts extraits de la MLT. Par exemple, si nous détectons *au niveau perceptuel* des mouvements spécifiques de la cible, nous pourrions n'extraire de la MLT que les concepts d'animaux. C'est bien sûr une question empirique.

<sup>17</sup> Le nombre d'exemplaires extraits de la MLT varie d'un modèle à l'autre. Certains modèles supposent que nous extrayons tous les exemplaires (Medin et Schaffer 1978).

les modèles fondés sur les prototypes font appel à différentes manières de calculer la ressemblance (voir le chapitre 4, section 2.3.).

La nature du processus de comparaison n'est pas spécifiée dans le schéma proposé. Non pas que cela n'ait pas d'importance. Bien au contraire. De nombreuses données empiriques, particulièrement sur la durée des performances de catégorisation, ne peuvent être utilisées que si les modèles de la catégorisation sont associés à des modèles relativement précis de ce processus de comparaison. Les modèles fondés sur les prototypes et les modèles fondés sur les exemplaires supposent différents processus de comparaison. Je propose ci-dessous un modèle du processus de comparaison pour chaque approche.

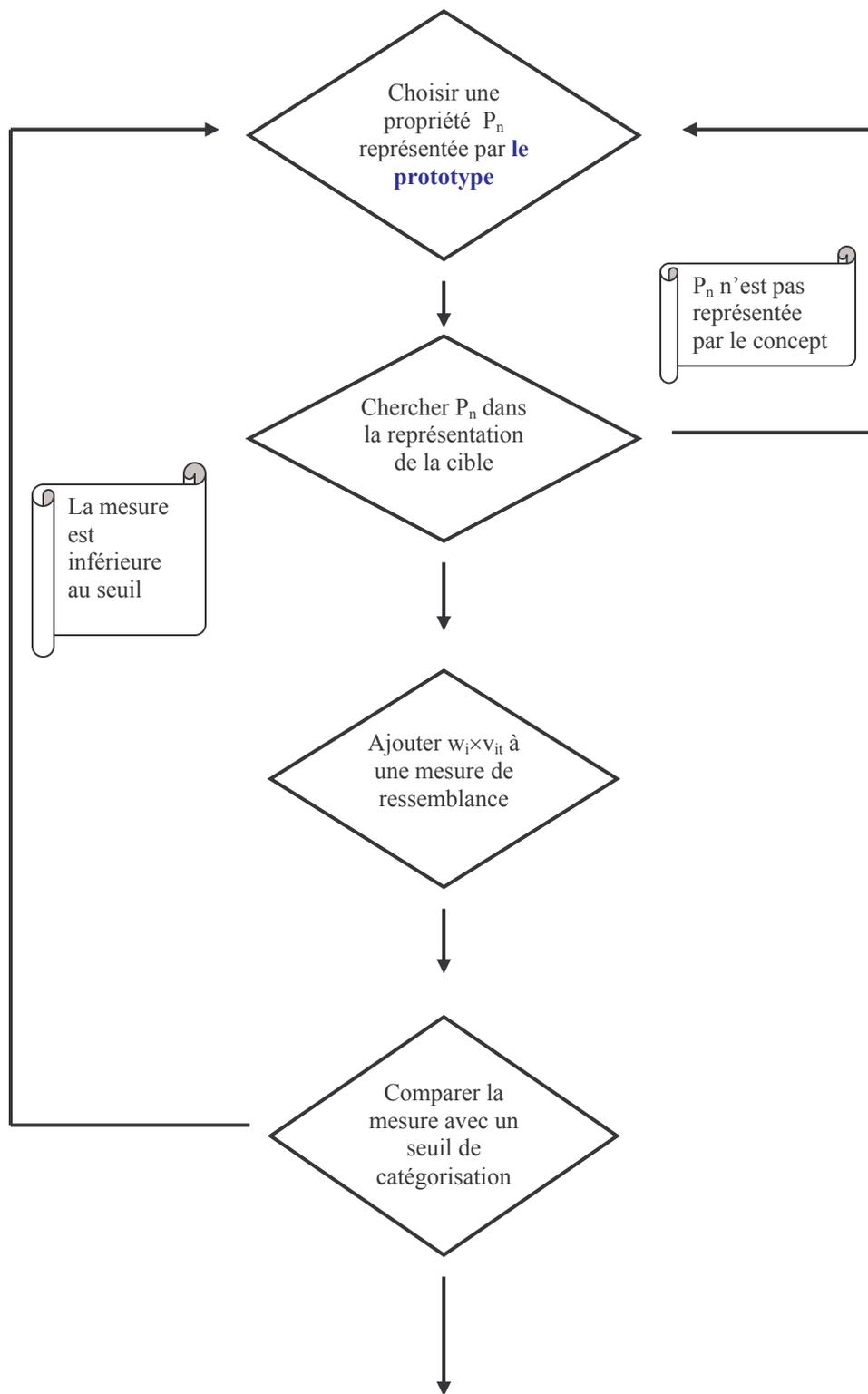


Schéma 5.6.

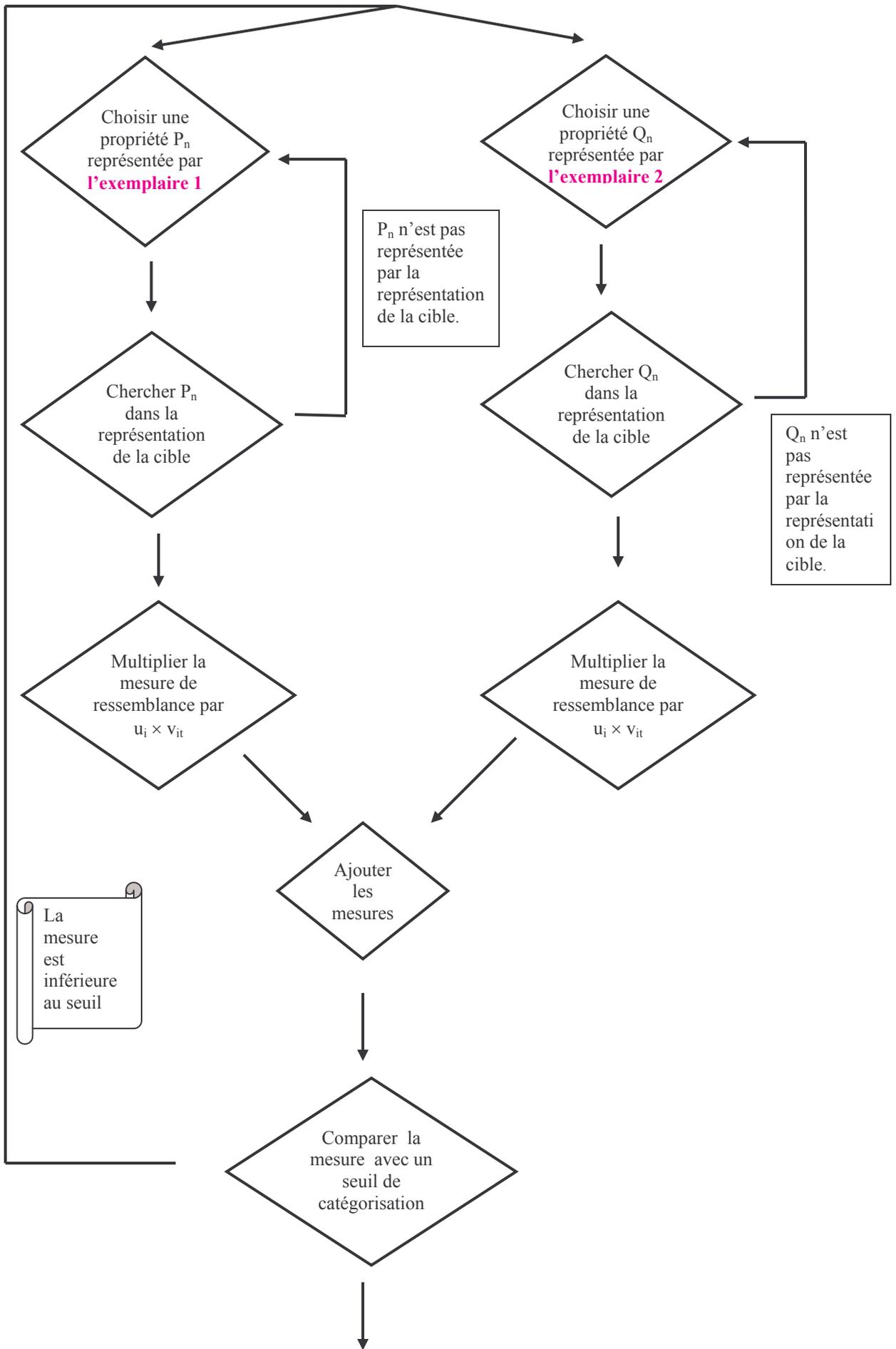


Schéma 5.7.

Ces deux modèles ont les propriétés suivantes. Dans les deux cas, la comparaison entre propriétés est *séquentielle* : les propriétés sont examinées les unes après les autres. En outre, la mesure de ressemblance est augmentée *progressivement* (par  $w_i \times v_{it}$  ou par  $u_i \times v_{it}$ , voir le chapitre 4, section 2.3.), et est constamment comparée à un seuil de catégorisation (éventuellement sensible au contexte).

Ces modèles permettent de faire des prédictions à propos de la durée de la catégorisation.<sup>18</sup> Dans ces deux modèles, une cible qui a de nombreuses propriétés en commun avec le prototype ou les exemplaires est catégorisée plus rapidement qu'une cible qui a peu de propriétés en commun avec eux. Car, dans ce cas, la plupart des propriétés qui sont représentées par le(s) concept(s) sont probablement possédées par la cible. En conséquence, le seuil de catégorisation devrait être atteint plus vite que si la cible n'a que quelques propriétés en commun avec le prototype. Ces modèles prédisent donc que les membres typiques d'une catégorie doivent être catégorisés plus rapidement que les objets atypiques.<sup>19</sup> C'est l'un des résultats les mieux établis de la psychologie des concepts (Rosch et Mervis 1975, voir la discussion dans Murphy et Lassaline 1997 et Murphy 2002, voir la section 4.1. de ce chapitre).

Soulignons que dans le modèle des exemplaires, la représentation de la cible est comparée avec de nombreuses représentations (les différents exemplaires d'une catégorie). Cette comparaison a lieu en parallèle. Au contraire, dans le modèle des prototypes, la représentation de la cible est comparée à une représentation unique, le prototype de la catégorie.

Revenons au schéma 5.5. Le dernier élément du schéma est *l'étape de décision*. Les mesures de ressemblance (correspondant aux différents concepts qui ont été extraits de la MLT durant l'étape de l'extraction) sont comparées à un seuil de catégorisation.

---

<sup>18</sup> Ce n'est pas le cas de tous les modèles. Certains modèles supposent que l'on calcule la ressemblance de la cible avec les concepts qui sont extraits de la MLT (Ross et Makin 1999). La cible est catégorisée dans la catégorie qui a le degré de ressemblance le plus élevé. Ce modèle ne permet pas de faire de prédiction sur la durée des performances de catégorisation.

<sup>19</sup> On utilise aussi la notion de ressemblance de famille. La ressemblance de famille d'un objet pour une catégorie donnée est (i) une fonction croissante du nombre de propriétés que cet objet a en commun avec les autres membres de la catégorie et (ii) une fonction décroissante du nombre de propriétés qu'il a en commun avec les objets qui n'appartiennent pas à la catégorie (Rosch et Mervis 1975 et la discussion dans Murphy 2002, 31 sq. ).

Dès qu'une mesure de ressemblance est supérieure à ce seuil, l'objet est catégorisé dans la catégorie correspondante. Si aucune mesure n'est supérieure à ce seuil, le système cognitif extrait d'autres concepts et, après un certain nombre d'extractions (éventuellement sensible au contexte), renonce à catégoriser la cible. Remarquons que c'est là une simplification : si nous échouons à catégoriser une cible, nous ne nous contentons pas de penser à d'autres catégories, nous examinons aussi de plus près la cible. Ainsi, nous ne nous contentons pas d'extraire d'autres concepts, nous modifions aussi, si possible, la représentation de la cible.

Notons finalement que ces modèles peuvent aussi rendre compte de nos jugements de typicalité. Les jugements de typicalité sont des jugements de la forme :

X est typique de la catégorie Y au degré z.

Les sujets déterminent souvent aisément si un objet est typique d'une catégorie donnée. Ils peuvent aussi classer des objets en fonction de leur typicalité et leur attribuer un degré de typicalité si on leur fournit une échelle (par exemple, une échelle selon laquelle 1 signifie atypique et 7 très typique). Par exemple, ils n'ont aucun problème à soutenir qu'un moineau est un oiseau très typique, tandis qu'un pingouin ne l'est pas. De nombreux partisans du PP et du PE supposent que les mécanismes qui expliquent nos performances de catégorisation expliquent aussi nos jugements de typicalité. La seule différence est que la mesure de ressemblance n'est pas comparée à un seuil de catégorisation. La typicalité d'un objet par rapport à la catégorie x est pour l'essentiel une fonction de la ressemblance de famille de x : un objet est jugé typique dans la mesure où il est semblable au concept (mais voir la note 4 du chapitre 4). Les modèles de catégorisation proposés ci-dessus (schémas 5.5., 5.6. et 5.7.) peuvent donc rendre compte de ces jugements de typicalité

## *2.2. Les modèles de la catégorisation fondés sur les théories*

Les partisans du PT rejettent habituellement ces modèles de la catégorisation (par exemple, Murphy et Medin 1985/2004, Rips 1989 – mais voir Heit 2001). Selon eux, la

catégorisation ne repose pas sur un calcul de ressemblance. Cependant, ils ne décrivent habituellement pas les mécanismes cognitifs qui réalisent notre compétence de catégorisation. Au lieu de cela, ils offrent une caractérisation intuitive de nos performances de catégorisation en termes de raisonnement et de processus inférentiels (par exemple, Gopnik et Metzoff 1997, 162-163). Par exemple, Rips décrit le processus cognitif qui sous-tend les performances de catégorisation de ses sujets comme suit (voir aussi l'exemple du plongeur ivre de Murphy et Medin) :

« Subjects' modal constructs in these examples – « it *must* be a bowl of rice » ; « they just *can't* be outside of that age group of *they're not even that thing* » ; (...) – indicate that *they are engaging in a form of reasoning that goes beyond simple distance comparison*. Subject C, for example, does not justify his choice on the grounds that 4,75 feet is closer to the size of a cereal box than to the size of a stop sign. Instead, he believes that something of that height simply can't be a stop sign (« it wouldn't make sense ») ; hence, by elimination, it must be a cereal box, which doesn't have these limitations on height. »<sup>20</sup> (Rips 1989, 37)

Les partisans du PT proposent souvent que les jugements de catégorisation reposent sur des inférences à la meilleure explication. Ainsi Rips écrit-il (1989, 52) :

« In most situations that call for categorizing, we confront some representation of an instance with our *knowledge* of the various categories it might belong to. *If the assumption that the instance is in one of these categories provides a reasonable explanation of the information we have about it and if this explanation is better than that provided by other candidate categories*, then we will infer that that instance is a member of the first category. »<sup>21</sup>

L'absence de modèles spécifiques de la catégorisation fondés sur des théories ne devrait pas trop nous surprendre. En effet, si nos performances de catégorisation résultent

---

<sup>20</sup> Je souligne.

<sup>21</sup> Je souligne.

d'inférences à la meilleure explication ou de processus semblables, il devrait être aussi difficile de caractériser les processus cognitifs qui réalisent notre compétence de catégorisation que n'importe quel autre aspect de notre raisonnement (Fodor 1983, Rips 1989, 52, 54, Smith et Sloman 1994, 378).

La catégorisation qui repose sur des théories peut être divisée en quatre étapes, à savoir la représentation de la cible, l'extraction du concept, l'étape du raisonnement et la décision de catégorisation. Les deux premières étapes sont semblables aux étapes correspondantes dans les modèles de la catégorisation qui font appel à des prototypes et à des exemplaires (voir le schéma 5.5.). L'étape de comparaison est remplacée par l'étape du raisonnement : nous nous engageons dans un raisonnement à propos de l'appartenance de la cible à la catégorie que nous sommes en train de considérer. Nous sommes supposés raisonner comme suit. Etant donné les propriétés de la cible et étant donné ce que nous savons à propos de la catégorie, est-il probable que la cible soit un membre de la catégorie ? Tandis que l'étape de comparaison est caractérisée en détail dans les modèles fondés sur des prototypes et les modèles fondés sur des exemplaires (voir les schémas 5.6. et 5.7.), cette étape du raisonnement n'a pas été décrite en détail. Le schéma 5.8. résume cette approche.

Au lieu d'un processus de raisonnement, on pourrait aussi suggérer que les théories spécifient des règles de catégorisation et que les processus de catégorisation appliquent ces règles (Smith et Sloman 1994, voir le chapitre 4, section 2.4.). Ces règles de catégorisation pourraient spécifier des conditions nécessaires et peut-être suffisantes d'appartenance. Par exemple, nos mini-théories à propos des rats pourraient spécifier que les rats sont nés d'autres rats (voir le chapitre 7, section 1.). Ainsi un concept théorique de rat contiendrait-il la règle : « un animal est un rat seulement si (et peut-être si et seulement si) il est né de rats ». Au lieu d'un processus de raisonnement, la troisième étape du processus de catégorisation consisterait à appliquer une règle (Smith et Sloman 1994, 377-378). Le schéma 6.8. peut être modifié en conséquence.

Finalement, contrairement aux modèles présentés ci-dessus, ce modèle n'est pas supposé expliquer aussi les jugements de typicalité. Car il n'inclut aucun calcul de ressemblance.

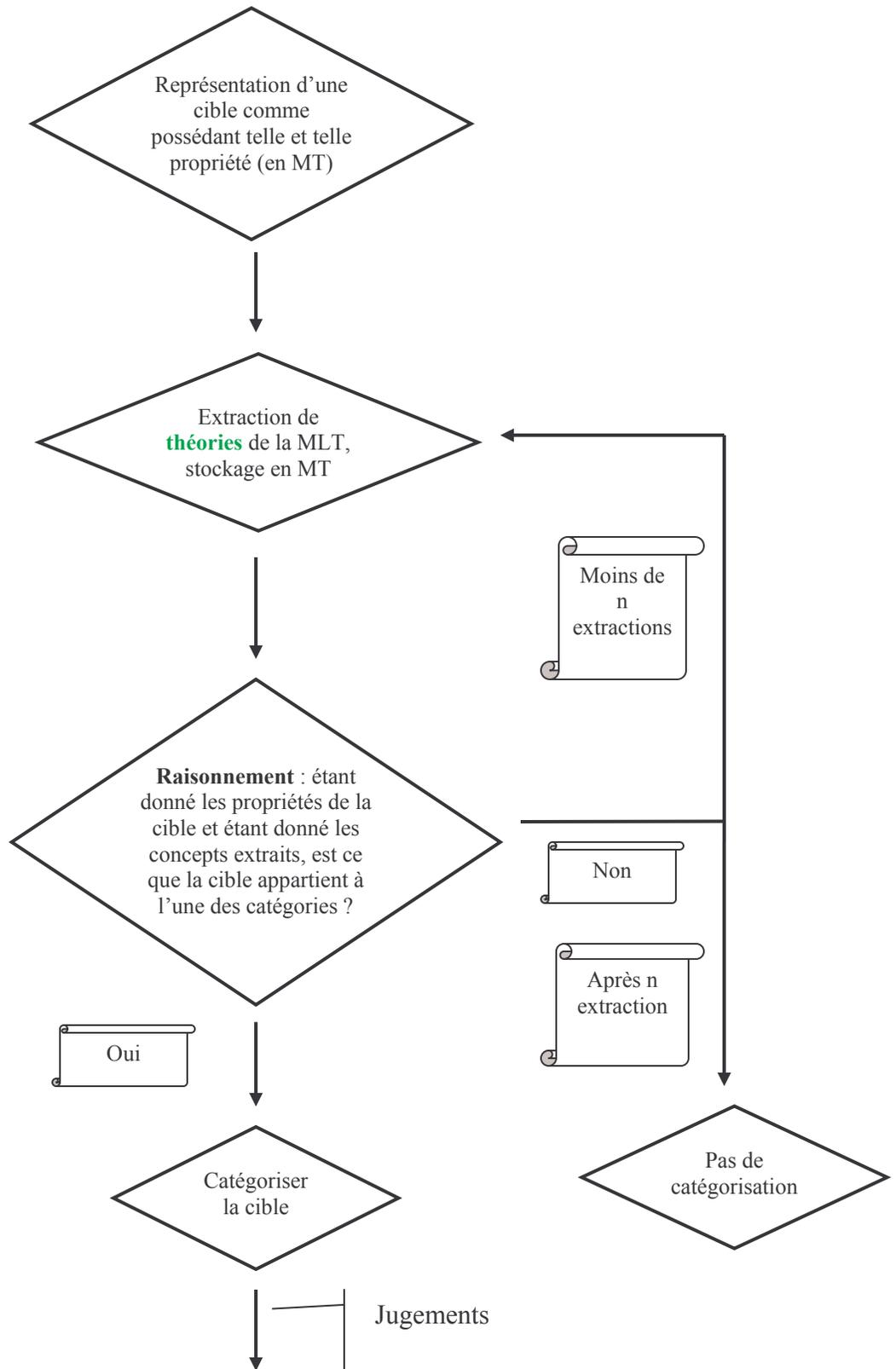


Schéma 5.8.

### 2.3. Conclusion

Dans la deuxième section de ce chapitre, j'ai proposé trois modèles de la catégorisation. Chacun d'eux repose sur un type spécifique de représentations, à savoir les prototypes, les exemplaires et les théories. Les deux premiers modèles reposent sur la ressemblance tandis que le troisième identifie la catégorisation à un processus de raisonnement. D'autres modèles fondés sur les prototypes et les exemplaires sont concevables. Les deux premiers modèles sont plus précis que le troisième, ce qui est en partie dû à la nature des processus cognitifs supposés. Ces modèles permettent d'utiliser les propriétés des performances de catégorisation – mises au jour dans des situations expérimentalement contrôlées – pour évaluer les théories de la catégorisation et des concepts.

## 3. Le processus de catégorisation et les théories : données empiriques

Dans cette section, je présente des données empiriques qui suggèrent que de nombreux jugements de catégorisation (à la fois des jugements d'inclusion et des jugements d'appartenance) résultent de processus qui impliquent des représentations théoriques.

### 3.1. Les dissociations entre les jugements de ressemblance et les jugements de catégorisation

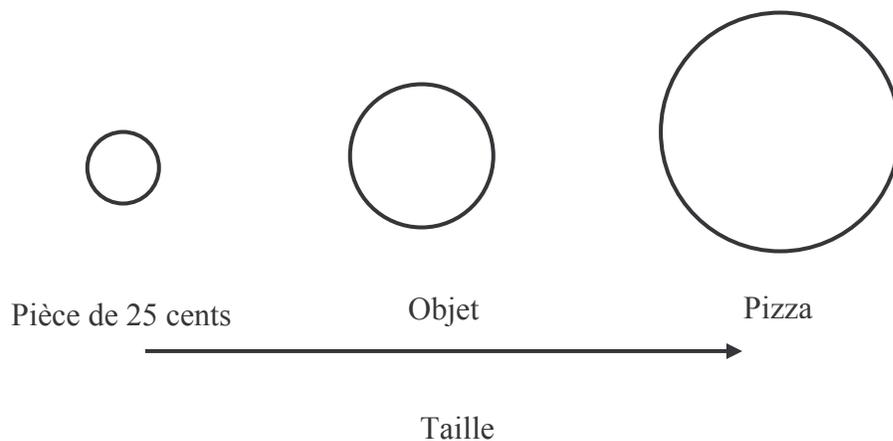
#### 3.1.1. Argument

Le premier type de données prétend montrer que les modèles de la catégorisation fondés sur des prototypes ou sur des exemplaires ne peuvent rendre compte de nombreux jugements de catégorisation. Selon ces modèles, les jugements de catégorisation reposent sur la ressemblance entre la cible (un objet ou une sous-classe) et le concept (un prototype ou un ensemble d'exemplaires). Cette idée de base implique qu'il devrait y avoir *une relation monotone entre la ressemblance avec le concept et la probabilité d'être catégorisé* (Hampton 1998) : plus une cible ressemble au concept, plus il est probable qu'elle soit catégorisée. Cependant, de nombreux partisans du PT soutiennent

que cette relation est souvent non monotone.<sup>22</sup> Je considère certaines données empiriques qui vont dans ce sens.

### 3.1.2. La pizza de Rips

L'expérience fameuse de la pizza, conduite par Lance Rips (1989), suggère que la relation entre les jugements de ressemblance et les jugements de catégorisation peut ne pas être monotone. Le paradigme expérimental est le suivant. Les sujets doivent imaginer un objet dont le diamètre est intermédiaire entre une pièce de 25 cents américaine et la plus petite pizza qu'ils aient vue. Aucune autre information n'est donnée aux sujets. La tâche consiste à juger (i) si l'objet ressemble davantage à la pizza ou à une pièce de 25 cents américaine et (ii) si cet objet est plus probablement une pizza ou une pièce de 25 cents.



**Schéma 5.9.**

Les résultats sont clairs : les sujets jugent que l'objet est aussi semblable à une pièce de 25 cents qu'à la plus petite pizza, tandis qu'ils jugent qu'il est plus probablement une pizza qu'une pièce de 25 cents (pour plus de détails, voir Rips 1989). Rips propose de manière convaincante que dans cette tâche de catégorisation, les sujets ont pris en compte la variabilité des deux catégories : puisque la taille d'une pièce de 25 cents est fixée par la loi, les sujets concluent de manière raisonnable que l'objet est plus probablement une pizza.

---

<sup>22</sup> On pourrait trouver cela évident : ne sait-on pas que les baleines ne sont pas des poissons ? Cependant, de telles connaissances sont sans doute stockées telles quelles en MLT. En conséquence, elles ne disent rien sur les processus de catégorisation.

On ne peut cependant conclure que *toutes* les catégorisations dépendent de ce type de processus. En effet, la dissociation entre les jugements de ressemblance et les jugements de catégorisation dépend de certains aspects particuliers du paradigme expérimental de Rips. Cherchant à répliquer ces résultats, Smith et Sloman (1994) ont montré que les jugements de catégorisation n'étaient pas dissociés des jugements de ressemblance quand on présentait aux sujets des descriptions plus longues de l'objet et quand on leur interdisait de raisonner verbalement pendant l'expérience (le raisonnement verbal est connu pour favoriser des modes de raisonnement fondés sur des règles).

Les résultats de Rips et ceux de Smith et Sloman montrent peut-être que les processus de catégorisation ne reposent pas sur la ressemblance dans des tâches verbales de catégorisation, contrairement aux tâches de catégorisation non verbales. *Cela vient à l'appui de l'hétérogénéité de la catégorisation*. En outre, puisque les modèles qui mettent en avant la ressemblance et ceux qui la rejettent font appel à des types différents de représentations (prototypes ou exemplaires d'une part, théories de l'autre), ces résultats suggèrent que différents types de concepts sont utilisés dans les processus de catégorisation. Dans des circonstances différentes, les individus ont donc recours à différents types de processus qui reposent probablement sur des types de concepts différents.<sup>23</sup>

Notons que certains partisans du PP ont soutenu que les résultats de Rips ne menacent pas de manière décisive les modèles de la catégorisation fondés sur les prototypes et ceux fondés sur les exemplaires (Hampton 1998, 2001). En effet, cette dissociation entre les jugements de ressemblance et les jugements de catégorisation ne montre pas nécessairement que le processus de catégorisation ne repose pas sur des prototypes ou des exemplaires. Car (i) certains modèles des prototypes et des exemplaires prédisent les jugements de catégorisation des sujets et (ii) les jugements de ressemblance ne reflètent pas le calcul de ressemblance qui, selon eux, a lieu dans les processus

---

<sup>23</sup> Smith et Sloman écrivent (1994, abstract) : « These findings are consistent with the hypothesis that categorization can be done in two ways, by *similarity* or by *rule*. An exclusive reliance on rule-based categorization seems to occur only when the description of the to-be-categorized object does not contain any features characteristic of any relevant category. » Mais voir la discussion dans Thibaut et al. 2002.

cognitifs. Concentrons-nous d'abord sur le premier point. Les jugements de catégorisation des sujets semblent reposer sur la variabilité de la taille des pizzas et des pièces américaines de 25 cents. Mais les modèles des prototypes peuvent représenter la variabilité d'une catégorie (voir le modèle de Smith et al. 1988 présenté dans le chapitre 3, section 2.2.) : si un prototype encode la fréquence de différentes propriétés pour un même attribut, par exemple de différentes tailles pour l'attribut taille, il est possible de déterminer si la taille des membres de la catégorie varie.<sup>24</sup> Selon les modèles standards, les exemplaires encodent implicitement la variabilité des catégories : en effet, il est possible de déterminer à partir des exemplaires de membres d'une catégorie si tous les membres de la catégorie ont la même propriété, par exemple la même taille. Dans certains modèles des prototypes et des exemplaires, ces représentations encodent donc la fréquence relative des propriétés. Si les concepts encodent, explicitement ou implicitement, la variabilité des catégories, on peut supposer que les dimensions peu variables sont pondérées davantage lors de la catégorisation (voir les pondérations des attributs dans le modèle de Smith et al. 1988, chapitre 3, section 2.2.). Ces modèles prédisent donc les jugements de catégorisation trouvés par Rips.

Mais qu'en est-il de la dissociation entre les jugements de catégorisation et les jugements de ressemblance ? Elle n'est pas nécessairement problématique pour les partisans des modèles de la catégorisation fondés sur les prototypes ou sur les exemplaires. Car les *jugements* de ressemblance ne reflètent pas nécessairement le *calcul* de ressemblance qui, selon eux, a lieu dans les processus de catégorisation qui font appel aux prototypes et aux exemplaires. Les jugements de ressemblance dépendent des dimensions qui sont contextuellement saillantes. La réponse à la question, « est-ce que ces deux objets se ressemblent ? », dépend de la dimension qui est saillante dans le contexte de discussion. Les instructions expérimentales utilisées par Rips suggèrent que la taille est la dimension pertinente pour le jugement de ressemblance, puisque c'est la seule information qui est fournie à propos de la cible. La variabilité des catégories n'est pas prise en compte dans le jugement de ressemblance dans la mesure où, dans l'expérience de Rips, elle n'est pas saillante. Cela explique la dissociation entre les jugements de ressemblance et les jugements de catégorisation.

---

<sup>24</sup> Sur la notion d'attribut, voir le chapitre 3, section 2.2..

Rips avait songé à une objection semblable.<sup>25</sup> Il répond que les descriptions verbales des sujets pendant l'expérience sont incompatibles avec cette hypothèse : les sujets utilisent des notions modales (nécessaire, possible) pour prendre leur décision de catégorisation. Ces données sont en effet suggestives. Elles viennent à l'appui de l'idée que la catégorisation repose parfois sur des concepts théoriques, puisque des concepts de ce type sont supposés encoder des informations modales. Remarquons cependant que l'argument ne repose plus sur la dissociation entre les jugements de catégorisation et les jugements de ressemblance. Ce qui compte maintenant contre les modèles de la catégorisation fondés sur les prototypes ou les exemplaires, ce sont les descriptions verbales avancées par les sujets. En outre, la verbalisation renforce les formes de cognition qui reposent sur des règles (voir la première objection ci-dessus). En conséquence, les expériences de Rips montrent au mieux que *certaines* performances de catégorisation ne reposent pas sur la ressemblance. Cela supporte l'hétérogénéité de la catégorisation, mais non le rejet intégral des modèles fondés sur la ressemblance, comme le défend Rips.<sup>26</sup>

En résumé, la dissociation entre les jugements de catégorisation et les jugements de ressemblance présentée par Rips ne réfute pas nécessairement les modèles fondés sur la ressemblance. Cependant, les descriptions verbales des sujets et les données de Smith et Sloman (1994) donnent à penser que les modèles fondés sur la ressemblance n'expliquent pas toutes les performances de catégorisation. Dans certains contextes (verbalisation, rareté des indices utilisables pour catégoriser...), on semble faire par défaut appel à des processus fondés sur des règles. Certaines performances sont donc mieux expliquées au moyen de modèles de la catégorisation qui dépendent des théories.

### 3.1.3. Les expériences de transformation de Rips

---

<sup>25</sup> Il écrit (Rips 1989, 36) : « Categorization subjects (...) used normed distance, with distance to the variable category normed by the variable category's standard deviation and distance to the fixed category normed by the fixed category's standard deviation. Similarity subjects, on the other hand, used unnormed distance to the two means. » Cela revient à dire que les sujets tiennent compte de la variabilité des catégories de pizzas et de pièces américaines de 25 cents lorsqu'ils calculent la ressemblance entre la cible et les prototypes/exemplaires de ces deux catégories dans le processus de catégorisation, mais pas lorsqu'ils émettent un jugement de ressemblance.

<sup>26</sup> En outre, les descriptions verbales que les sujets font de leurs propres processus cognitifs ne sont pas un type de données parfaitement fiable (voir Nisbett et Wilson 1977, mais voir Malt 1989).

Rips (1989) présente une autre dissociation entre les jugements de catégorisation et les jugements de ressemblance.<sup>27</sup> Les sujets lisent des histoires qui décrivent la transformation physique d'animaux dans des circonstances variées. Par exemple, un animal qui ressemblait à un oiseau ressemble à un insecte suite à un accident industriel (condition accident). Dans une autre version (condition essence), un animal qui ressemblait à un oiseau ressemble à un insecte à cause d'un processus endogène.<sup>28</sup> L'histoire suivante correspond à la condition accident (Rips 1989, 39) :

« There was an animal called a sorp which, when it was fully grown, was like other sorps, having a diet which consisted of seeds and berries found on the ground or on plants. The sorp had two wings, two legs, and lived in nests high in the branches of a tree. Its nest was composed of twigs and other fibrous plant material. This sorp was covered with bluish-gray feathers.

The sorp's nest happened to be not too far from a place where hazardous chemicals were buried. The chemicals contaminated the vegetation that the sorp ate, and as time went by it gradually began to change. The sorp shed its feathers and sprouted a new set of wings composed of a transparent membrane. The sorp abandoned its nest, developed a brittle iridescent outer shell, and grew two more pairs of legs. At the tip of each of the sorp's six legs, an adhesive pad was formed so that it was able to hold onto smooth surfaces ; for example, the sorp learned to take shelter during rainstorms by clinging upside down to the undersides of tree leaves. The sorp eventually sustained itself entirely on the nectar of flowers.

Eventually, this sorp mated with a normal female sorp one spring. The female laid the fertilized eggs in her nest and incubated them for three weeks. After that time normal young sorps broke out of their shells. »

On demande aux sujets (i) si la créature ressemble davantage à un oiseau ou à un insecte et (ii) s'il est plus probable qu'elle soit un oiseau ou un insecte. Dans la condition accident, les sujets disent qu'elle ressemble davantage à un insecte, même si elle est plus

---

<sup>27</sup> Ces expériences sont semblables aux expériences de Keil (1989, chap. 8).

<sup>28</sup> Certaines histoires impliquent aussi des artéfacts.

probablement un oiseau. Dans la condition essence, les sujets soutiennent qu'avant le changement, la créature était plus probablement un insecte, même si elle ressemblait davantage à un oiseau. Rips conclut que les jugements de catégorisation ne sont pas une fonction des jugements de ressemblance. Il souligne aussi le fait que les jugements de catégorisation reposent sur la nature des processus causaux impliqués (croissance endogène contre causes externes).

Ces résultats n'ont cependant pas été répliqués par Estes et Hampton (résultats décrits dans Hampton 2001). Estes et Hampton n'étaient pas satisfaits par certains aspects de l'expérience conduite par Rips. En particulier, Rips utilise une procédure intra-sujets (« within-subjects design ») : les sujets doivent répondre successivement aux deux questions, ce qui suggère pour des raisons pragmatiques que des réponses différentes sont attendues. Avec une procédure entre-sujets (« between-subjects design »), Estes et Hampton montrent que les jugements de catégorisation ne sont pas dissociés des jugements de ressemblance.

En outre, Hampton (2001) soulignent que ces expériences de transformation ne sont pas les plus adéquates pour étudier les processus cognitifs qui réalisent notre compétence de catégorisation. Etant donné le caractère extraordinaire de ces transformations, il n'est pas certain que les décisions de catégorisation des sujets dans ces tâches reposent sur leurs processus de catégorisation habituels.

Il n'est donc pas certain que les résultats de Rips viennent à l'appui des modèles de la catégorisation fondés sur les théories.

### *3.2. Création de catégories et connaissances causales*

Les psychologues de la catégorisation utilisent de nombreux paradigmes expérimentaux. Mais ceux-ci peuvent être classés en deux types :

- *Classification.* Les sujets connaissent les catégories utilisées par l'expérimentateur. L'expérience porte sur la manière dont ils décident d'inclure un objet dans ces catégories. La connaissance de ces catégories peut avoir diverses sources. Par exemple, elle peut résulter de la familiarité des

sujets avec les catégories utilisées dans l'expérience : c'est le cas, par exemple, de la catégorie des oiseaux (par exemple, Rips 1989). Elle peut aussi résulter de l'apprentissage de ces catégories pendant l'expérience (par exemple, Rosch et Mervis 1975).

- *Création de catégories.* On présente aux sujets des objets. L'expérience porte sur la manière dont les sujets décident de grouper ensemble ces objets et de former ainsi des catégories. Cette procédure est supposée révéler les croyances que possèdent les sujets sur la nature des catégories. Dans certains cas, la classification est *libre*. On présente aux sujets des objets et on leur demande de grouper ceux qui vont ensemble (par exemple, Ahn et Dennis 2001, présenté ci-dessous). Dans d'autres cas, l'expérimentateur a au préalable défini les catégories auxquelles appartiennent les objets et corrige les décisions de classification des sujets (par exemple, Conant et Trabasso 1964).

Les expériences présentées dans la section précédente (section 3.1.) reposent toutes sur le paradigme de classification. Mais l'étude expérimentale de la compétence de catégorisation repose aussi sur la seconde procédure expérimentale. Les psychologues supposent en effet que la manière dont nous formons des catégories éclaire la manière dont nous classifions des objets dans les catégories que nous connaissons déjà. Car la création de catégories révèle nos croyances sur la nature des catégories. Ces croyances révèlent quelles connaissances sont encodées par nos concepts et utilisées par les processus cognitifs impliqués dans la catégorisation.

La procédure de création de catégories a donc été utilisée pour défendre les approches de la catégorisation fondées sur les théories. L'idée de base est la suivante. La création d'une catégorie à partir d'objets présentés aux sujets ne peut être expliquée en termes statistiques (ce que soutient le PP) ou par référence à la ressemblance avec certains membres de la catégorie (ce que soutient le PE). Au contraire, les sujets prennent en compte les relations causales entre les propriétés des objets (ainsi que d'autres connaissances théoriques) quand on leur demande de former une catégorie (voir, par exemple, Murphy et Medin 1985/2004, Medin et Wattenmaker 1987, Keil et al. 1998).

Cela montre que les sujets croient que les catégories ne reposent pas seulement sur la ressemblance entre leurs membres, mais aussi sur des relations causales. Cela est prédit si nos concepts sont des concepts théoriques.

Considérons, par exemple, les expériences récentes de Ahn et Dennis (2001). Ils avaient l'ambition de montrer que par défaut, nous attribuons davantage d'importance aux propriétés qui sont causalement centrales<sup>29</sup> pour grouper des objets ensemble. Cela n'est pas prédit par les modèles fondés sur les prototypes ni par les modèles fondés sur les exemplaires.

La procédure de la seconde expérience est la suivante. On présente aux sujets une cible qui est caractérisée par trois propriétés causalement reliées, A, B et C : A cause B qui cause C. Deux objets sont décrits. L'objet  $\Phi$  a en commun la propriété causalement centrale avec la cible et possède deux propriétés spécifiques D et E : A cause D qui cause E. L'objet  $\Psi$  a en commun avec la cible les deux propriétés qui ne sont pas causalement centrales et possède une propriété spécifique causalement centrale F : F cause B qui cause C. Par exemple, Ahn et Dennis utilisent les descriptions suivantes :

#### **Target**

Judy has a cough which was caused by Virus XB12.

Judy has a severe headache which was caused by the cough.

#### **Object $\Phi$**

Chris has a stuffed nose which was caused by Virus XB12.

Chris has shortness in breath which was caused by the stuffed nose.

#### **Object $\Psi$**

Mary has a cough which was caused by Virus Zomba.

Mary has a severe headache which was caused by the cough.

---

<sup>29</sup> C'est-à-dire, les propriétés dont nous croyons qu'elles causent la possession d'autres propriétés. Par exemple, *être incurvé* est une propriété centrale de l'un de nos concepts de boomerangs. Car nombre de personnes croient (à tort) qu'*être incurvé* explique causalement pourquoi les boomerangs reviennent quand ils ont été lancés.

On pose aux sujets la questions suivante : quel objet devrait être catégorisé avec la cible ? 75% des sujets répondent que l'objet  $\Phi$  doit être catégorisé avec la cible. Cette expérience révèle donc que nous supposons que les catégories sont organisées autour de propriétés causales centrales : nous supposons que les membres d'une même catégorie ont en commun un certain nombre de propriétés causales qui déterminent causalement la possession d'autres propriétés.<sup>30</sup> Si l'on admet que la création de catégories révèle la nature des processus de catégorisation, ces résultats montrent que les modèles de la catégorisation fondés sur les prototypes et les modèles de la catégorisation fondés sur les exemplaires n'expliquent pas toutes les propriétés de nos performances de catégorisation.<sup>31</sup> Au contraire, des connaissances théoriques, semblables à celles que les concepts théoriques sont supposés encoder, résultent de la création de catégories et sont donc probablement utilisées dans la catégorisation.

Notons toutefois qu'il n'est pas certain que ces résultats valent généralement pour toutes nos catégorisations. En particulier, la création de catégories sur la base des informations causales résultent peut-être de la saillance des informations causales dans les stimuli utilisés. Ainsi, les résultats de Ahn et Dennis caractérisent peut-être plutôt un contexte particulier de catégorisation, lié à la saillance de la causalité.

Remarquons cependant que ces données ne viennent pas particulièrement à l'appui de l'idée selon laquelle la catégorisation repose sur l'inférence à la meilleure explication ou sur des raisonnements modaux. Car elles sont compatibles avec l'idée que notre compétence de catégorisation est réalisée par un processus de comparaison qui implique un calcul de ressemblance (Ahn et Dennis 2001). On pourrait en effet imaginer que le processus de catégorisation est le suivant. Les propriétés qui sont représentées par

---

<sup>30</sup> Keil et al. écrivent pareillement (1998, 111, voir aussi Medin et Wattenmaker 1987) : « When the features available for learning novel instances of natural categories are interrelated, people use knowledge of these interrelations to guide learning a new category without having to attend to raw feature frequencies. But when the features are not interrelated, people must resort to strategies using only the raw statistical frequencies of features. »

<sup>31</sup> On pourrait trouver étrange de comparer des « objets » du genre moineau etc. à des « objets » du genre symptômes médicaux. Mais le PP, le PT et le PT font l'hypothèse qu'un même type de représentations mentales dénote ces différents types de catégories (voir la discussion de la seconde interprétation de l'hypothèse de l'hétérogénéité, chapitre 2, 2.3. b.).

un concept et les propriétés qui sont attribuées à la cible sont comparées. On pondère les propriétés qui sont à la fois représentées par le concept et attribuées à la cible et on les additionne pour produire une mesure de ressemblance. La décision de catégorisation dépend de cette mesure. Cette description est fidèle aux modèles courants de la catégorisation qui sont fondés sur la ressemblance (voir les schémas 5.5., 5.6. et 5.7.). Cependant, les pondérations qui sont attribuées aux propriétés qui sont communes au concept et à la cible sont fonction de leur importance causale. Les propriétés causalement centrales sont davantage pondérées. Ainsi la catégorisation requiert-elle des connaissances que ni les exemplaires ni les prototypes ne sont supposées encoder.<sup>32</sup> Mais le processus cognitif n'est pas une inférence à la meilleure explication.

En résumé, les expériences de Ahn suggèrent que dans certains contextes, nos performances de catégorisation reposent par défaut sur des concepts théoriques, bien qu'elles ne montrent pas – et ne cherchaient pas à montrer – que les modèles de la catégorisation fondés sur les théories, qui rejettent les calculs de ressemblance, décrivent correctement certaines de nos performances de catégorisation. En d'autres termes, les expériences de Ahn et Dennis montrent que des théories sont utilisées dans la catégorisation, mais elles ne montrent pas comment elles sont utilisées. Elles ne supportent donc pas les modèles de la catégorisation qui font appel à des processus de raisonnement plutôt qu'à des processus de comparaison.<sup>33</sup>

### 3.3. Conclusion

Les données empiriques présentées ci-dessus, ainsi que d'autres données dans la littérature psychologique, suggèrent que dans certains contextes, nous utilisons par défaut certaines connaissances théoriques quand nous prenons une décision de catégorisation. Cela n'est expliqué ni par le PP ni par le PE. En conséquence, nous devons admettre l'existence de concepts théoriques. Certaines données, quoique plus ambiguës, suggèrent

---

<sup>32</sup> Ahn et Dennis écrivent ainsi (2001, 105) : « It is possible (...) to shift the meaning of « similar » to be more inclusive of deeper features (...). It may be, then, that categorization does depend on this expanded notion of similarity. Similarity judgments based on all of the available information could allow people to use the proper weighting that is used in categorization. This weighting is based on, but is probably not limited to, (...) the causal status of features. »

<sup>33</sup> Cependant, nous avons vu ci-dessus que les descriptions verbales des sujets suggèrent que nous utilisons bien, dans certaines circonstances, des considérations modales et des inférences à la meilleure explication pour prendre des décisions de catégorisation.

aussi que ces concepts théoriques sont utilisés dans des inférences à la meilleure explication et dans des processus analogues qui ne reposent pas sur la ressemblance.

#### **4. Les processus de catégorisation, les prototypes et les exemplaires : données empiriques**

##### *4.1. A-t-on besoin de postuler des prototypes pour expliquer la compétence de catégorisation ?*

De nombreuses propriétés des performances de catégorisation suggèrent que de nombreuses performances de catégorisation reposent sur des calculs de ressemblance. Je compare dans cette section les explications de ces phénomènes qui sont fournies par le PP et le PE.<sup>34</sup> Cette comparaison montre que le pouvoir explicatif du PE est supérieur au pouvoir explicatif du PP. Cela pose donc le problème suivant : a-t-on besoin de postuler des prototypes pour expliquer nos performances de catégorisation ?

- **La corrélation entre la typicalité et d'autres mesures de la catégorisation**

Cette corrélation a été établie dans les années 70 par Rosch, Mervis, Rips etc. Elle a joué un rôle crucial dans le succès du PP (Rips et al. 1973, Smith et al. 1974, Rosch 1978, Rosch et Mervis 1975).

Ce phénomène bien connu est le suivant. Pour une classe donnée d'objets, les membres qui sont jugés typiques par un individu sont catégorisés plus rapidement et avec moins d'erreurs par cet individu que les membres moins typiques (par exemple, Rips et al. 1973). On trouve ce phénomène aussi bien pour des stimuli qui sont présentés visuellement (on montre par exemple aux sujets l'image d'un moineau et on leur demande si c'est un oiseau) et verbalement (on pose aux sujets la question suivante : « est-ce qu'un moineau est un oiseau ? »).

➤ *L'explication proposée par le PP.* La typicalité d'un objet est une fonction de sa ressemblance avec le prototype de la catégorie concernée. La ressemblance d'un objet

---

<sup>34</sup> Je me concentre pour l'essentiel sur la psychologie expérimentale. Pour des données en neurosciences cognitives, voir Knowlton et Squire 1995, Ashby et al. 1998, Reber et al. 1998, Smith et al. 1998, Knowlton 1999, Smith et Minda 2000b.

typique avec le prototype dépasse donc le seuil de catégorisation plus rapidement que la ressemblance d'un objet atypique. Par ailleurs, les erreurs du système cognitif (défaut d'attention...) affectent moins la décision de catégorisation. Les modèles 5.5. et 5.6. prédisent donc que les objets jugés typiques sont catégorisés plus rapidement et avec moins d'erreurs que les objets jugés atypiques.

➤ *L'explication proposée par le PE.* La typicalité d'un objet est une fonction de sa ressemblance avec les exemplaires stockés en mémoire à long terme. La ressemblance d'un objet jugé typique avec les exemplaires dépasse donc le seuil de catégorisation plus rapidement que la ressemblance d'un objet jugé atypique. Par conséquent, le seuil de catégorisation est atteint plus rapidement pour les objets typiques et les erreurs du système cognitif affectent moins leur catégorisation. Les modèles 5.5. et 5.7. prédisent donc que les objets jugés typiques sont catégorisés plus rapidement et avec moins d'erreurs que les objets jugés atypiques.

- **La typicalité et le vague**

L'appartenance de certains objets à certaines catégories n'est ni clairement positive ni clairement négative : leur appartenance est *vague*. C'est, par exemple, le cas de la catégorie des tas, comme le montre la question « Est-ce que trois grains de sable forment un tas ? ». Dans ces circonstances, les sujets ne s'accordent pas les uns avec les autres sur l'appartenance de ces objets aux catégories concernées. Leurs jugements varient aussi d'une occasion à l'autre. McCloskey et Glucksberg (1978) ont montré que la typicalité prédit les désaccords entre les jugements de catégorisation émis par différentes personnes et par la même personne à différents moments : leurs jugements divergent pour les objets qui ne sont ni très typiques ni très atypiques.

➤ *L'explication proposée par le PP.* Les objets qui ne sont ni typiques ni atypiques ne sont ni très semblables au prototype de la catégorie concernée ni très différents de lui. Ils ont donc un degré de ressemblance avec le prototype qui est proche du seuil de catégorisation. Cela explique les jugements d'appartenance vague. Par ailleurs, leur

catégorisation est fortement affectée par de nombreux facteurs (attention etc.). Cela explique la variabilité des jugements.

➤ *L'explication proposée par le PE.* Les objets qui ne sont ni typiques ni atypiques ressemblent seulement à quelques exemplaires stockés en MLT. Cela explique les jugements d'appartenance vague. Par ailleurs, leur catégorisation est fortement affectée par de nombreux facteurs (attention etc.).<sup>35</sup> Cela explique la variabilité des jugements.

- **L'effet prototype**

Cet effet a été décrit pour la première fois par Posner et Keele (1968) et a été souvent reproduit (Posner et Keele 1970, Homa et al. 1981, Knowlton et Squire 1993, Reber et al. 1998). Le paradigme expérimental utilisé pour mettre au jour cet effet est le suivant. On présente aux sujets des objets qui appartiennent à une catégorie donnée et les sujets doivent former un concept de cette catégorie. Dans la phase de test, les sujets doivent décider si des objets appartiennent ou non à la catégorie (voir le paradigme de classification dans la section 3.2.). La catégorie qui doit être apprise par les sujets est structurée autour d'un membre typique : elle est composée d'un membre typique et d'autres membres qui sont plus ou moins semblables à ce membre. Par exemple, dans Posner et Keele (1968, 1970), les sujets doivent apprendre des catégories dont les membres sont des patterns de points. Ces patterns sont obtenus par déformation d'un pattern de point original – qui est le membre typique. Durant la phase d'apprentissage, on présente aux sujets des membres de la catégorie. Le membre typique n'est pas présenté. Pendant la phase de test, les sujets doivent décider si des objets cibles appartiennent ou non à la catégorie. L'effet prototype est triple :

- Si la phase de test suit immédiatement la phase d'apprentissage, le membre typique est catégorisé plus précisément (c'est-à-dire, avec moins d'erreurs) que des membres nouveaux (c'est-à-dire, non présentés durant la phase

---

<sup>35</sup> Cette explication ne suppose-t-elle pas que l'agent ait encodé des exemplaires d'objets qui soient représentatifs de leur catégorie, par exemple des chiens représentatifs ? Ce n'est pas le cas. Car il s'agit d'expliquer les jugements que fait l'agent lui-même. Si l'agent a encodé des exemplaires non représentatifs, ces jugements sur l'appartenance vague seront simplement *sui generis*.

d'apprentissage), bien qu'il soit catégorisé moins précisément que les objets utilisés pendant la phase d'apprentissage (Posner et Keele 1968).

- La précision de la catégorisation des nouveaux objets est une fonction de leur typicalité.
- Si la phase de test ne suit pas immédiatement la phase d'apprentissage, le membre typique est catégorisé plus précisément que les membres utilisés pendant la phase d'apprentissage (Posner et Keele 1970, Homa et al. 1981).

On pourrait se demander si ce résultat n'est pas propre à la perception visuelle de bas niveau. Les stimuli sont en effet des objets qui n'ont pas de signification sémantique : ils sont au contraire définis uniquement par des propriétés perceptuelles. Dans quelle mesure peut-on généraliser l'effet prototype à l'ensemble de la catégorisation, en particulier à la catégorisation dans des catégories qui ne sont pas définies seulement par des propriétés perceptuelles ? La question est pertinente. Notons tout d'abord que les autres effets décrits dans cette section se manifestent aussi bien avec des catégories visuelles, comme une catégorie de patterns de points, que pour des catégories non perceptuelles. Pourquoi l'effet prototype serait-il une exception ? Par ailleurs, on trouve aussi l'effet prototype quand on utilise des catégories qui ne sont pas dépourvues de signification, par exemple avec des visages (Cabeza et al. 1999). Cependant, il se peut que l'effet ne soit pas reproduit si l'on utilise des stimuli verbaux plutôt que visuels.

➤ *L'explication proposée par le PP.* Pendant la phase d'apprentissage, les sujets extraient un prototype à partir des objets qui leur sont présentés. Supposons que les objets présentés soient représentatifs de la catégorie. Supposons aussi que ce prototype soit extrait à partir de tous les objets présentés. Supposons enfin qu'il encode la valeur moyenne des propriétés qui caractérisent les objets présentés. Par conséquent, le membre typique correspond davantage au prototype que les nouveaux objets. L'hypothèse selon laquelle nous catégorisons sur la base de la ressemblance à un prototype prédit alors les deux premiers éléments de l'effet prototype. Quand les souvenirs des objets présentés lors de la phase d'apprentissage s'effacent (par effacement des traces mnésiques –

« memory decay »), ces objets sont catégorisés comme s'ils étaient nouveaux. Cela explique le troisième élément.

➤ *L'explication proposée par le PE.* Supposons que nous créons un exemplaire pour chaque objet présenté. Supposons aussi que les objets présentés sont représentatifs. La typicalité des objets est alors corrélée à leur ressemblance aux exemplaires. L'hypothèse que nous catégorisons sur la base de la ressemblance à des ensembles d'exemplaires prédit donc les deux premières propriétés de l'effet prototype. Cependant, la troisième propriété n'est pas directement prédite par le PE : si nos décisions de catégorisation reposent sur des exemplaires, les objets présentés pendant la phase d'apprentissage devraient être classifiés plus facilement que les nouveaux objets, y compris l'objet le plus typique.

Cependant, on a montré que dans certaines conditions, cet effet est prédit par le PE (Hintzman et Ludlam 1980). L'objet typique qui est présenté pendant la phase de test ressemble à de nombreux exemplaires. Au contraire, un objet qui a été utilisé pendant la phase d'apprentissage peut ne ressembler qu'à peu d'exemplaires. En conséquence, malgré sa nouveauté, le degré de ressemblance de l'objet typique peut être plus élevé que celui d'un objet qui a été présenté dans la phase d'apprentissage. En outre, les exemplaires en MLT peuvent en partie s'effacer : ils peuvent perdre une partie des connaissances qu'ils encodaient à l'origine. En conséquence, quand un objet qui a été présenté pendant la phase d'apprentissage est présenté de nouveau dans la phase de test, il peut ne pas correspondre exactement à l'exemplaire en MLT.

- **L'effet exemplaire**

On présente aux sujets des membres d'une catégorie pendant la phase d'apprentissage. Pendant la phase de test, ils doivent catégoriser des objets dans cette catégorie. L'effet exemplaire est double.

- Tout d'abord, les objets qui sont également typiques d'une même catégorie sont catégorisés d'autant plus précisément qu'ils ressemblent davantage aux objets présentés pendant la phase d'apprentissage (Whittlesea 1987).

- Deuxièmement, les objets qui sont très semblables à certains objets présentés pendant la phase d'apprentissage peuvent être catégorisés plus précisément que des objets plus typiques (Medin et Schaffer 1978).

➤ *L'explication proposée par le PP.* Cet effet n'est pas une conséquence immédiate du PP. Il peut être expliqué par l'hypothèse *ad hoc* que la phase de test suit immédiatement la phase d'apprentissage et que les objets qui sont présentés pendant la phase d'apprentissage sont toujours stockés en mémoire. Cette explication prédit que l'effet exemplaire doit disparaître si la phase de test ne suit pas immédiatement la phase d'apprentissage.

➤ *L'explication proposée par le PE.* Ce phénomène est une conséquence immédiate du PE, puisque la mesure de ressemblance est non linéaire (voir l'explication donnée dans le chapitre 4, section 2.3.).

- **L'effet de la variabilité et l'effet de la corrélation**

La variabilité des catégories peut être contrôlée expérimentalement au moyen de catégories d'objets artificiels. Pareillement, l'existence de corrélations entre les propriétés possédées par les membres d'une catégorie peut être contrôlée. Il est donc possible d'étudier comment la variabilité d'une catégorie et la corrélation entre les propriétés affectent nos performances de catégorisation.

- *L'effet de la variabilité* est le suivant : les sujets prennent en compte la variabilité d'une catégorie pour prendre une décision de catégorisation (Fried et Holyak 1984, Rips 1989, ci-dessus, section 3.1.).
- *L'effet de la corrélation* est le suivant : les sujets prennent en compte les corrélations entre propriétés pour prendre une décision de catégorisation (Malt et Smith 1984).

➤ *L'explication proposée par le PP.* Ces deux phénomènes ne sont pas prédits par la plupart des modèles des prototypes. Cependant, dans certains modèles, les prototypes

représentent la variabilité des catégories (par exemple, Smith et al. 1988, et ci-dessus, section 3.1.). Pareillement, les prototypes peuvent en principe encoder les corrélations de propriétés (Hampton 2001). Cependant, cela peut conduire à une explosion combinatoire : même si l'esprit ne tenait compte que d'un petit nombre de propriétés des membres des catégories, l'esprit devrait encoder un très grand nombre de corrélations binaires de propriétés, de corrélations ternaires etc.

➤ *L'explication proposée par le PE.* La variabilité des catégories et les corrélations entre propriétés ne sont pas explicitement représentées par les exemplaires. Cependant, selon le PE, l'esprit encode implicitement ce type de connaissances (voir le chapitre 4, section 4.1.). Les corrélations entre les propriétés et la variabilité de la catégorie dénotée peuvent être calculées à partir des propriétés qui sont explicitement représentées par les exemplaires. La capacité de prendre en compte la variabilité d'une catégorie et les corrélations entre propriétés ne produisent par ailleurs aucune explosion combinatoire. Car l'esprit ne les encode pas explicitement.

Ci-dessus, j'ai comparé les explications qui ont été proposées par le PP et le PE pour rendre compte d'importantes propriétés de la catégorisation. Quelle conclusion faut-il tirer de cette comparaison ? Le PE explique les données que le PP peut expliquer. En outre, il explique aisément l'effet exemplaire, l'effet de la corrélation et l'effet de la variabilité qui sont problématiques pour le PP.

Généralement, une donnée empirique qui est supposée venir à l'appui du PP peut souvent être expliquée par le PE. Car les connaissances qui sont supposées être explicitement encodées par les prototypes peuvent être implicitement encodées par les ensembles d'exemplaires. Elles pourraient être extraites quand le besoin s'en fait sentir (voir le chapitre 4, section 4.1.).<sup>36</sup> En conséquence, peu de données empiriques montrent de manière décisive que les individus utilisent des prototypes dans les tâches de catégorisation.

---

<sup>36</sup> Rappelons toutefois qu'un ensemble d'exemplaires encode exactement la même information qu'un prototype seulement si l'on possède un exemplaire pour chaque membre rencontré (voir Barsalou 1990 et le chapitre 4, section 4.2.).

En résumé, les données présentées ci-dessus montrent que les performances de catégorisation impliquent un calcul de ressemblance. Les modèles de la catégorisation fondés sur les exemplaires semblent mieux expliquer ces données que les modèles fondés sur les prototypes. Doit-on conclure que l'explication de la catégorisation nous autorise à postuler des théories (voir section 3.) et des exemplaires, mais non des prototypes ?

#### 4.2. *Un argument pour l'existence de prototypes*

Supposons que nous dénotions une catégorie C au moyen d'un ensemble d'exemplaires S. Au début, nous n'avons pas de prototype de C. Quand nous catégorisons un objet i dans C, nous comparons i aux membres de S (ou éventuellement un sous-ensemble, dépendant du contexte, de S). Supposons aussi que nous utilisions régulièrement la variabilité de C et la corrélation entre les propriétés parmi les membres de C pour décider si de nouveaux objets appartiennent à C. Selon le PE, nous devons *extraire* ces connaissances de S à *chaque fois* que nous en avons besoin. Car elles sont encodées *implicitement* dans S. Cependant, selon certaines versions du PP, un prototype les encoderait *explicitement* (Smith et al. 1988). Si nous avons un tel prototype de C, nous n'aurions pas besoin d'extraire ces connaissances de S à chaque fois que nous en avons besoin. Au lieu d'extraire nos connaissances sur la variabilité et les corrélations de propriétés, il semble plus économique de stocker un prototype en MLT. En conséquence, l'idée d'économie cognitive suggère que nous possédons un prototype en sus de l'ensemble d'exemplaires.

On peut trouver cet argument spéculatif. Et il l'est. L'idée d'économie cognitive n'est pas très claire et son importance est discutable. Par ailleurs, il est difficile de mesurer ce qui est plus coûteux, l'extraction répétée de connaissance stockées implicitement ou le stockage en mémoire à long terme d'un type supplémentaire de représentations mentales. En conséquence, on pourrait rejeter un argument qui fait appel à cette idée. Cependant, l'idée présentée ci-dessus a été développée par un programme de recherches empiriques, appelé l'explication de l'apprentissage de catégories fondée sur la réminiscence (« reminding-based account of category learning ») ou l'explication de l'apprentissage de catégories fondée sur la comparaison (« comparison-based account of

learning ») (voir, par exemple, Ross et al. 1990, Spalding et Ross 1994, et la recension dans Ross et Makin 1999<sup>37</sup>). L'idée de base de ce programme est la suivante :

« The current proposal is that *reminders that occur during the learning of a category lead to generalizations*. During the early training of a category, it is assumed that people often make their categorization judgments by similarity to previous instances. That is, a current instance reminds them of an earlier instance, so, in the absence of any other available information, they assign it to the same category as the earlier instance (e.g., Brooks 1978). The main extension of this proposal beyond exemplar theories is the hypothesis that reminding results in the explicit comparison of the instances and the storage of those aspects that the instances have in common. Later instances may be categorized by reminding the learner of earlier instances or by the access and use of these generalizations from earlier reminders. Depending on the relation between instances in the category, the use of these generalizations may lead to still further abstractions. Thus, both general and specific information is stored. » (Ross et al. 1990, 464)

Ces psychologues ont amassé de nombreuses données empiriques qui viennent à l'appui de cette hypothèse. Je décris ci-dessous la première expérience de Ross et al. 1990. Elle présente des données qui ne sont explicables simplement ni par le PP ni par le PE. Ces données suggèrent au contraire (i) que, pour catégoriser de nouveaux objets, les sujets font appel à des représentations d'individus singuliers qui ont été présentés pendant la phase d'apprentissage et (ii) qu'ils comparent les objets qui sont jugés appartenir à cette classe pour abstraire des connaissances générales à propos de la catégorie et qu'ils utilisent ces connaissances dans leurs processus cognitifs.

La procédure expérimentale est la suivante. Dans la phase d'apprentissage, les sujets doivent apprendre certains faits à propos de quatre individus ( $A_1$ ,  $A_2$ ,  $B_1$ ,  $B_2$ ) qui appartiennent à deux catégories, A et B. Ces quatre individus sont caractérisés par trois propriétés. La phase de test est double. Tout d'abord, on présente aux sujets la description verbale d'un individu. Il est caractérisé par trois propriétés. Deux propriétés sont appelées

---

<sup>37</sup> Homa et al. 1981 et Medin et Edelson 1988 ont développé des idées semblables.

« les propriétés critiques » (« critical features ») : chacune est thématiquement reliée à une propriété d'un individu présenté dans la phase d'apprentissage, disons  $A_1$ . Le dernier trait est thématiquement relié à une propriété de l'autre individu de la même catégorie,  $A_2$ . En conséquence, l'individu décrit dans la phase de test a plus de chances d'être catégorisé comme un A *sur la base de sa ressemblance avec  $A_1$* . Deuxièmement, quatre propriétés sont ensuite présentées aux sujets. Deux d'entre elles sont neutres (« filler objects »). Les sujets doivent décider quelle propriété a le plus de chances d'être possédée par un nouveau membre de A. Pour illustrer cette procédure, je présente ci-dessous un exemple des stimuli utilisés dans cette expérience (Ross et al. 1990, 446).

<i>Phase d'apprentissage</i>		
Instance	Shirley	Julia
Aime	Glaces	Westerns
Achète	Clous	Maillots de bain
Catégorie	A	A

<i>Test 1 (catégorisation) : les sujets doivent catégoriser un individu décrit par 3 propriétés</i>	
Aime	Sorbets
Achète	Bois
Achète	Serviette de plage

<i>Test 2 : les sujets doivent classer les quatre propriétés ci-dessous (c et d sont neutres)</i>	
a	Achète un marteau
b	Achète des lunettes de soleil
c	Achète des chaussettes
d	Achète du pain

### **Schéma 5.10.**

L'idée est la suivante. L'individu décrit dans la première phase est supposée être catégorisé comme un A en vertu de sa ressemblance avec Shirley : il possède deux propriétés qui sont thématiquement en rapport avec les propriétés de Shirley. Selon

l'explication de la catégorisation fondée sur la réminiscence, à l'occasion de cette catégorisation, les sujets forment la généralisation que les membres de A ont des activités qui sont en rapport avec le bois ou la menuiserie. En conséquence, dans le second test, ils sont supposés conclure qu'un membre quelconque de A a plus de chances d'acheter un marteau que des lunettes de soleil. Les résultats confirment cette prédiction.

Ni le PP ni le PE ne peuvent séparément expliquer ce phénomène. Si les sujets extraient des connaissances sur la fréquence des propriétés à partir d'exemplaires de A ou s'ils forment un prototype à partir des deux premiers objets et s'ils le modifient sur la base du troisième, ils n'ont aucune raison de supposer qu'un membre quelconque de A a plus de chances d'acheter un marteau que des lunettes de soleil. Car les sujets ont vu trois membres de A. Au total, deux propriétés sont en rapport avec la charpenterie (acheter des clous et acheter du bois) et deux propriétés sont en rapport avec la plage (acheter un maillot de bain et acheter une serviette de plage).

Ces données viennent à l'appui de l'argument avancé ci-dessus.<sup>38</sup> Quand nous utilisons des exemplaires dans une tâche cognitive, nous tendons à former des prototypes.<sup>39</sup> Nous formons des prototypes en sus des exemplaires. Notons toutefois que selon Ross et collègues, les pondérations des propriétés encodées par les prototypes n'encodent pas uniquement la fréquence des propriétés, mais aussi leur utilisation dans les catégorisations.

#### *4.3. Catégorisation et neuropsychologie*

En outre, d'autres programmes de recherche montrent que nous possédons et utilisons aussi bien des prototypes que des ensembles d'exemplaires. Knowlton a montré que notre compétence de catégorisation et notre capacité de reconnaître des individus peuvent être dissociées (Knowlton et Squire 1993, Knowlton 1997, 1999 voir le chapitre 2, section 4.,

---

<sup>38</sup> Ross et al. 1990 ont invalidé plusieurs explications de ce phénomène qui feraient exclusivement appel aux prototypes ou aux exemplaires. Ces résultats ont été reproduits avec une procédure un peu différente dans Spalding et Ross 1994.

<sup>39</sup> Ross et collègues proposent parfois une interprétation de ces résultats qui est un peu différente. Comme moi, ils concluent que les individus utilisent à la fois des connaissances générales et des connaissances singulières. Spalding et Ross (1994) semblent parfois soutenir que les connaissances générales ne sont pas explicitement représentées, mais sont encodées par des pondérations qui mettent en valeur certaines connaissances qui sont encodées par les exemplaires au dépens d'autres connaissances (voir aussi Medin et Edelson 1988, 79-80).

mais voir Nosofsky et Zaki 1998). L'expérience est semblable aux expériences de Posner et Keele (1968, 1970). Pendant la phase d'apprentissage, on présente aux sujets des patterns de points. Ces patterns ont été obtenus par distorsion d'un pattern original. Ce pattern est le membre de la catégorie qui est le plus semblable au prototype de la catégorie. La phase de test a lieu une semaine plus tard. On présente aux sujets des patterns de points. Ils doivent les catégoriser et reconnaître les patterns présentés lors de l'entraînement (tâche de reconnaissance ancien/nouveau). La principale différence avec les expériences de Posner et Keele est que les sujets souffrent d'amnésie. Il existe différentes formes d'amnésie. Les sujets de Knowlton sont incapables de se souvenir des objets rencontrés (voir la discussion de Knowlton 1999). Les troubles amnésiques de ces patients résultent, selon Knowlton, d'un déficit de *stockage de l'information*.<sup>40</sup> Autrement dit, ces patients ne sont pas seulement incapables d'extraire de leur mémoire les souvenirs qu'ils ont stockés (déficit de l'extraction des connaissances). Ils sont incapables de stocker de nouveaux souvenirs.

Les résultats sont les suivants. Les patients amnésiques sont incapables de reconnaître les patterns de points utilisés dans la phase d'apprentissage. Cela est compatible avec leurs troubles amnésiques : ils ne sont pas capables de reconnaître ces patterns de points, parce qu'ils n'ont pas stocké de souvenirs singuliers des objets durant la phase d'apprentissage. Ces patients ne peuvent donc pas avoir utilisé des exemplaires dans cette tâche de catégorisation. Cependant, ils catégorisent les patterns de points comme le font les sujets normaux : le prototype et les patterns de points légèrement déformés sont reconnus plus facilement que les patterns de points fortement déformés (voir la discussion de l'effet prototype). Ils ont donc probablement abstrait un prototype des objets présentés lors de l'apprentissage. Knowlton conclut donc (1997, 236, je souligne) :

*« In the real world, category learning is undoubtedly not uniform, and the processes and knowledge representations involved probably differ depending on the exemplars and the training procedure used. In some cases, when individual examples are not memorable or distinctive, subjects may form a prototype*

---

<sup>40</sup> L'hippocampe de ces patients est affecté. L'hippocampe est impliqué dans la formation des souvenirs.

representation. However, when the individual examples are distinctive, then subjects may use exemplar-based knowledge for classification. Under some circumstances, subjects may be using declarative knowledge of examples to make classification judgments. »

#### *4.4. Les effets de facilitation*

Comme nous l'avons vu au chapitre 4 (section 3.), de nombreux psychologues ont douté que l'on puisse distinguer expérimentalement les modèles inspirés par le PE et ceux inspirés par le PP (Malt 1989, Barsalou 1990). En réponse, Malt a soutenu que les psychologues devraient utiliser de nouvelles mesures des performances de catégorisation (Malt 1989). L'idée est la suivante. Supposons que le PE soit vrai. Les individus utilisent donc des représentations singulières pour catégoriser les objets. Quand ils ont utilisé une représentation singulière, cette représentation devrait être disponible pour d'autres processus cognitifs. En conséquence, une autre tâche qui dépend de cette représentation devrait être facilitée. Au contraire, supposons que le PP soit correct. Les individus utilisent des prototypes pour catégoriser les objets. Leurs performances dans une tâche qui dépendrait d'une représentation singulière ne devraient donc pas être facilitées par une catégorisation antérieure.

La première expérience de Malt consiste à tester la validité de cette hypothèse. La procédure est la suivante. Des paires d'animaux sont créées : les membres de ces paires se ressemblent physiquement (par exemple, un jaguar et un lion constituent une paire). Les animaux sont divisés en deux groupes, A et B (par exemple, le lion et le jaguar appartient à A). Pendant la phase d'apprentissage, on présente aux sujets un membre de chaque paire (par exemple, le jaguar). Ils doivent apprendre la catégorie à laquelle il appartient (A). La phase de test est la suivante. On demande aux sujets de catégoriser les animaux, y compris ceux utilisés dans la phase d'apprentissage. Les membres d'une même paire (par exemple, le jaguar et le lion) sont présentés à la suite l'un de l'autre, le nouvel objet (le lion) étant présenté en premier. On demande explicitement aux sujets de catégoriser le nouvel animal (le lion) en se remémorant l'animal qui est le plus semblable à la cible (le jaguar) et en plaçant la cible (le lion) dans la catégorie de cet animal. En fait, on demande aux sujets de catégoriser au moyen d'exemplaires. L'idée est que dans ces

conditions, la catégorisation d'un nouvel objet (par exemple, le lion) repose sur le membre associé (le jaguar) dans la paire définie par Malt, puisque ces deux animaux sont très semblables. Puisque cette représentation est activée, la tâche de catégorisation suivante (la catégorisation du jaguar) devrait être plus facile.

Pour détecter un effet de facilitation, une condition contrôle est mise en place. On demande aux sujets de décider si un animal nouveau (le lion) est plus grand qu'un cocker. Cela ne requiert pas d'extraire une représentation singulière de l'objet présenté dans la phase d'apprentissage (l'exemplaire du jaguar). Après chaque objet, on donne aux sujets la tâche de catégorisation avec les objets présentés dans la phase d'apprentissage (le jaguar). Cette condition fournit une référence afin de mesurer d'éventuels effets de facilitation dans la première tâche. On donne enfin aux sujets une tâche de reconnaissance : ils doivent décider si les objets ont été présentés pendant la phase d'apprentissage.

Les résultats montrent que la procédure expérimentale de Malt détecte bien l'extraction d'exemplaires pendant le processus de catégorisation. Tout d'abord, les performances de reconnaissance sont très bonnes (4-5% d'erreurs), ce qui montre que les sujets ont encodé les objets présentés pendant l'apprentissage et sont capables d'extraire leurs représentations. Ensuite, on trouve un effet significatif de facilitation : les sujets sont significativement plus rapides pour catégoriser les objets après avoir dû catégoriser l'objet associé dans les paires définies par Malt plutôt qu'après une question sur la taille des animaux. Cela vient à l'appui de l'hypothèse de Malt : sa procédure expérimentale détecte adéquatement l'extraction d'exemplaires durant le processus de catégorisation.<sup>41</sup>

Puisque ce paradigme est sensible à l'utilisation d'exemplaires dans les tâches de catégorisation, il peut être utilisé pour déterminer si les sujets utilisent des prototypes et/ou des exemplaires dans les tâches de catégorisation. L'expérience 3 porte sur cette question (voir aussi l'expérience 6). La procédure est semblable à celle de l'expérience 1. La principale différence est que dans l'expérience 3, on ne donne aux sujets aucune instruction explicite de catégorisation. Ils peuvent utiliser n'importe quelle stratégie de catégorisation. Pendant la phase de test, on demande aux sujets qui prennent part à la

---

<sup>41</sup> Malt a invalidé plusieurs explications différentes de ses résultats (expériences 2 et 5). En particulier, elle montre dans l'expérience 5 que l'extraction d'un prototype ne produirait aucun effet de facilitation.

condition principale de catégoriser des animaux nouveaux et anciens comme dans l'expérience 1. On leur demande aussi de décrire leurs stratégies de catégorisation. La condition contrôle est semblable à celle de l'expérience 1.

Les résultats montrent que les sujets utilisent *plusieurs stratégies*. Certaines décisions de catégorisation reposent sur des exemplaires, tandis que d'autres font appel à des prototypes. En effet, en agrégeant les données, Malt obtient un effet de facilitation. Mais cet effet n'est pas significatif (contrairement à l'expérience 1). Cela suggère que sans instructions explicites, tous les sujets ne font pas appel à des exemplaires.<sup>42</sup>

Les descriptions que font les sujets de leurs stratégies expérimentales viennent à l'appui de cette interprétation.<sup>43</sup> Malt écrit ainsi (1989, 546-547) :

« 3 said they used only general features if the category in classifying the new exemplars. Nine said they used only similarity to old exemplars, and 8 said that they used a mixture of category features and similarity to old exemplars. If reports accurately reflect the strategies used, then the data are composed of responses involving several different decision processes. The mixture of strategies will, of course, reduce the amount of priming overall compared with Experiment 1, where subjects were presumably all using an exemplar retrieval strategy. »

De nouveau, ces résultats montrent que nous devons postuler des prototypes en sus des exemplaires.

Les expériences recensées ci-dessus proviennent de plusieurs champs disciplinaires et utilisent des méthodes différentes. Cela vient à l'appui de mon hypothèse : nous avons des prototypes *et* des exemplaires.

#### 4.5. Conditions déclenchantes

Dans le chapitre 4, j'ai soutenu qu'une même compétence pouvait être réalisée par plusieurs processus (section 3.). Ce chapitre défend l'hétérogénéité de la catégorisation :

---

<sup>42</sup> Ce résultat est reproduit avec différents stimuli dans l'expérience 6.

<sup>43</sup> L'expérience 6 montre que ces descriptions reflètent sans doute les stratégies réellement utilisées. En effet, Malt trouve une corrélation entre la stratégie décrite (usage d'un prototype ou d'exemplaires) et l'existence d'un effet de facilitation (Malt 1989, 552, mais voir aussi Medin et Edelson 1988, 73 où les descriptions fournies par les sujets sont erronées ou non informatives).

notre compétence de catégorisation repose sur plusieurs processus. Cela soulève naturellement la question suivante : *comment ces processus sont-ils articulés ?*

*Cette question empirique est encore sans réponse.* La théorie duelle de la cognition et la théorie pluraliste des processus cognitifs que je défends n'en sont qu'à leur début. Les psychologues ont davantage cherché à montrer l'existence de plusieurs processus qu'à étudier leur articulation (voir Smith et Sloman 1994, Hahn et Chater 1998, Smith et al. 1998, Stanovich et West 2000, 2003, Haidt 2001).

On peut cependant faire les suggestions suivantes (voir en particulier Malt 1989, Allen et Brooks 1991, Smith et Sloman 1994, Sloman 1998, Smith et al. 1998).

➤ Ces processus peuvent parfois fonctionner en parallèle et produire des outputs différents (Allen et Brooks 1991, Smith et al. 1998, Greene et al. 2001). Les données sur la durée des performances viennent à l'appui de cette idée. Par exemple, dans les expériences rapportées dans Allen et Brooks 1991, les sujets doivent catégoriser des objets nouveaux dans deux catégories. Certains objets satisfont la règle qui caractérise l'une des catégories, mais sont très semblables aux objets qui appartiennent à l'autre catégorie. Ces objets sont catégorisés plus lentement que des objets non ambigus (voir aussi Smith et al. 1998). Cela suggère que les sujets doivent choisir entre les outputs de deux processus.<sup>44</sup> Dans d'autres cas, les outputs peuvent être intégrés. Smith et Sloman (1994) soutiennent ainsi que même quand une règle peut être employée pour catégoriser un objet (par exemple, la règle « si un objet a les propriétés P, Q, et R, c'est un z »), les sujets prennent aussi en compte les propriétés typiques dans leurs décisions de catégorisation (voir aussi Hampton 1995).

➤ Certaines circonstances peuvent déclencher particulièrement l'un des processus. Les autres processus pourraient être déclenchables intentionnellement dans ces circonstances. Les expériences d'Allen et Brooks (1991) suggèrent que quand les stimuli sont visuels, les processus qui reposent sur la ressemblance sont automatiquement déclenchés. Dans ces conditions, les processus qui ne reposent pas sur la ressemblance peuvent être intentionnellement déclenchés. Par ailleurs, Smith et Sloman (1994) ont montré que les raisonnements exprimés linguistiquement favorisent le processus de catégorisation qui ne repose pas sur la ressemblance.

---

<sup>44</sup> Le même argument a été avancé pour les jugements moraux (Greene et al. 2001).

#### *4.6. Conclusion*

Les phénomènes présentés dans cette section sont difficilement explicables par le PT. En conséquence, les modèles de la catégorisation fondés sur les théories sont incomplets. Il faut postuler d'autres types de processus. Au premier abord, ces phénomènes ne semblent requérir que des exemplaires : postuler des prototypes semble superflu. Cependant, des considérations de plausibilité et des données empiriques qui proviennent de sources variées montrent que nous devons postuler des prototypes en sus des exemplaires. En conséquence, nos performances de catégorisation reposent sur des théories, sur des exemplaires et sur des prototypes (et peut-être sur d'autres types de représentations). Bien sûr, cela vient fortement à l'appui de la prémisse 2 de l'ANE.

### **5. Une objection**

#### *5.1. L'hétérogénéité de la catégorisation*

L'introduction d'entités théoriques n'est légitime que si cela permet d'expliquer au mieux des phénomènes scientifiquement importants. Les chapitres 5-7 portent sur cette question. En particulier, je soutiens que la meilleure explication de nombreuses propriétés des compétences cognitives supérieures repose sur l'hypothèse de l'hétérogénéité, c'est-à-dire l'hypothèse que nous avons plusieurs types de concepts qui ont peu en commun.

La section 3 de ce chapitre a montré que certaines propriétés des catégorisations sont naturellement expliquées au moyen de théories. Certaines catégorisations reposent sur les connaissances que les théories sont supposées encoder. Ni le PP ni le PE n'expliquent cela. La section 4 a montré que d'autres propriétés sont naturellement expliquées au moyen de prototypes et d'exemplaires. De nombreuses données suggèrent que nous utilisons des connaissances sur les individus et des connaissances statistiques pour catégoriser. En conséquence, certaines catégorisations reposent sur les connaissances que les exemplaires et les prototypes sont supposés encoder. Le PT ne peut expliquer cela.

Ces résultats semblent donc venir à l'appui de l'hétérogénéité de la catégorisation. Si nous postulons différents types de concepts, à savoir des prototypes, des exemplaires

et des théories (l'hypothèse de l'hétérogénéité) et si nous postulons qu'ils figurent dans différents processus de catégorisation (l'hétérogénéité de la catégorisation), nous expliquons l'ensemble des données présentées dans ce chapitre.

### 5.2. *Le problème*

Mais est-ce vraiment la meilleure explication de ces données ? J'ai suggéré qu'aucun des trois paradigmes n'était capable à lui seul d'expliquer ces données. Mais qu'en est-il d'autres paradigmes ?

Il semble que les théories composites des concepts soient à même d'expliquer les données présentées dans ce chapitre. Selon ces théories (voir le chapitre 2, section 3.2.), les concepts sont divisés en parties qui encodent des types différents de connaissances. Supposons qu'une partie d'un concept encode des connaissances sur les individus particuliers, qu'une autre encode des connaissances statistiques à propos des propriétés possédées par la plupart des membres de la catégorie dénotée et qu'une troisième encode certaines connaissances théoriques (voir Gelman 1990, 2001). Selon cette théorie composite des concepts, *un unique concept* encode exactement les connaissances qui sont encodées, selon moi, par *trois* représentations mentales différentes, un prototype, un ensemble d'exemplaires et une théorie. En conséquence, les données empiriques qui semblent montrer que nos processus cognitifs utilisent ces trois types de connaissances viennent tout autant à l'appui de cette théorie composite des concepts. Par exemple, le fait que nous utilisons ces trois types de connaissances quand nous catégorisons est également conforme à cette théorie composite des concepts et à l'HH.

En outre, les partisans de cette théorie pourraient soutenir que selon la maxime de parcimonie, les scientifiques devraient chercher à minimiser le nombre d'entités admises dans leur ontologie. Cette idée semble favoriser la théorie composite des concepts présentée ci-dessus, puisqu'elle ne postule qu'un seul type de concepts au lieu de trois. Cet argument peut être résumé comme suit :

### **L'argument de la parcimonie pour les théories composites des concepts**

1. Certaines théories composites des concepts supposent qu'un concept encode exactement les connaissances qu'un prototype, qu'un ensemble d'exemplaires et qu'une théorie coréférentiels encoderaient.
2. L'étude d'une compétence cognitive supérieure, par exemple la catégorisation, montre que cette compétence repose sur les types de connaissances que les prototypes, les exemplaires et les théories sont supposés encoder.
3. La prémisse 2 vient également à l'appui de l'HH et de la théorie composite des concepts (d'après 1).
4. La théorie composite des concepts est plus parcimonieuse.
5. Donc, on devrait admettre cette théorie (de 3 et 4).

Si cet argument est correct, les données empiriques à propos des compétences cognitives supérieures ne viennent pas à l'appui de l'hypothèse de l'hétérogénéité et du rejet de l'hypothèse EN.

#### *1.3. Réponse*

Les théories composites des concepts sont les principales rivales de l'hypothèse de l'hétérogénéité. Car la plupart des données qui viennent à l'appui de l'HH viennent également à l'appui de certaines théories composites. Je ne vais pas répondre maintenant à cette objection. Le chapitre 8 reviendra sur cette question.

## **6. Conclusion**

Ce chapitre a pour objet de défendre la prémisse 2 de l'ANE : il est justifié d'admettre dans notre ontologie psychologique des exemplaires, des prototypes et des théories. J'ai fait la recension de phénomènes psychologiques importants qui ont été mis au jour par l'étude empirique de nos performances de catégorisation. La catégorisation est la capacité de produire des jugements d'appartenance et des jugements d'inclusion. Plusieurs modèles de cette compétence ont été proposés en psychologie. J'en ai présenté trois : ils font respectivement appel aux prototypes, aux exemplaires et aux théories. J'ai ensuite

présenté des données empiriques qui montrent que certaines catégorisations reposent sur des théories, d'autres sur des exemplaires et finalement d'autres sur des prototypes. Cela est compatible avec la prémisse 2 de l'ANE et avec l'hypothèse de l'hétérogénéité.

## Chapitre 6

### La composition des concepts

---

Ce chapitre est dans la lignée du chapitre 5. Après avoir considéré en détail la psychologie de la catégorisation, je me tourne maintenant vers une autre compétence cognitive supérieure, à savoir la composition des concepts, afin de défendre la prémisse 2 de l'argument de la nécessité explicative (ANE). Je soutiens que l'étude empirique de la composition des concepts vient à l'appui de la prémisse 2 de l'ANE : il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer la composition des concepts. En effet, je défends l'idée qu'il existe un processus cognitif de composition des concepts qui implique les prototypes, les exemplaires *et* les théories.

Malgré les nombreux désaccords qui existent à propos des concepts, philosophes et psychologues s'accordent pour soutenir que, quelle que soit la nature des concepts, ils se composent. Depuis l'article fondateur d'Osherson et Smith (1981), les psychologues cognitifs ont beaucoup travaillé sur ce sujet. En effet, Osherson et Smith postulèrent explicitement que *les théories des concepts devaient être à même de rendre compte de la composition des concepts*. Sur cette base, ils avancèrent l'idée que les théories des prototypes, associées avec la théorie des ensembles vagues de Zadeh, étaient inadéquates pour rendre compte de nos jugements de catégorisation dans des catégories représentées par des concepts complexes. En psychologie, ce fut l'une des premières attaques contre le PP, qui était alors l'approche dominante. La plupart des psychologues acceptèrent le postulat de Osherson et Smith (par exemple, Medin et Shoben 1988, Wisniewski 1996, 434, Wisniewski et Love 1998, 177, Murphy 1988, 2002, Costello et Keane 2000, 2001). Il en résultat un effort prolongé pour rendre compte de la composition des concepts. Malgré l'intensité de cet effort, les psychologues ne sont nullement arrivés à un consensus : comme pour la catégorisation, le désaccord règne parmi les psychologues des concepts et des fonctions cognitives supérieures sur la nature du ou des processus cognitifs qui produisent des concepts complexes (voir la recension des recherches menées dans ce domaine par Murphy 2002, chapitre 12).

En outre, la composition des concepts est devenue un sujet brûlant en philosophie et entre philosophes de la psychologie et psychologues. En particulier, Jerry Fodor a soutenu (i) que le fait que les concepts se composent est la seule chose que nous savons pour sûr à propos des concepts et (ii) qu'aucune théorie psychologique des concepts n'est compatible avec ce fait (voir, par exemple, Fodor 1998).<sup>1</sup> En raison des arguments de Fodor, la composition des concepts est devenue un sujet obligé en philosophie de la psychologie (voir Laurence et Margolis 1998, Cowie 1999, chapitre 6, Gärdenfors 2000, chapitre 4, Prinz 2002, chapitre 12). La recherche psychologique sur les concepts a aussi été affectée par ces débats (voir, par exemple, l'échange entre James Hampton et Fodor dans *Mind and Language*, 15, 2-3, 2000, et plus récemment Connolly et al.<sup>2</sup>).

Dans ce chapitre, je ne recense pas la recherche psychologique sur les concepts depuis les travaux de Osherson et Smith 1981 (voir les recensions de Rips 1995, Hampton 1997b et Murphy 2002). Je ne décris pas non plus les débats philosophiques sur la question (Fodor et Lepore 1996, Fodor 1998, 1999a, b, Villanueva 1998, Machery 2002, Récanati 2002 etc.). Je propose de décrire l'un des mécanismes cognitifs impliqués dans la composition des concepts. Je soutiens que les données empiriques viennent à l'appui de la thèse suivante : *la composition des concepts résulte entre autres<sup>3</sup> d'un mécanisme cognitif qui utilise à la fois des prototypes, des exemplaires et des théories*. En conséquence, les données expérimentales à propos de la composition des concepts viennent à l'appui de la prémisse 2 de l'ANE.

Dans la première section, je caractérise la compétence de composition et je décris l'objet d'une théorie de la composition des concepts. Dans la seconde section, je soutiens qu'il existe un processus de composition des concepts<sup>4</sup> qui implique des prototypes. Dans la troisième section, je montre qu'il faut postuler non seulement des prototypes, mais aussi des concepts théoriques. Dans la quatrième section, je décris et j'explique un autre

---

<sup>1</sup> Il soutient en fait qu'à l'exception de sa propre théorie atomistique des concepts, aucune théorie des concepts, aussi bien en philosophie qu'en psychologie, n'est compatible avec ce fait.

<sup>2</sup> Connolly, A. C., Fodor, J. A., Gleitman, L. R., et Gleitman, H., 2003, « Why stereotypes don't even make good defaults », manuscrit.

<sup>3</sup> Je soutiens ailleurs que la composition des concepts repose sur plusieurs mécanismes, la plupart étant des processus cognitifs simples, rapides et liés à des contextes spécifiques (voir Machery, E., « Simple heuristics for concept composition », manuscrit).

<sup>4</sup> Dans ce qui suit, j'utilise par simplicité l'expression « la composition des concepts ». Il faut entendre *un* processus de composition des concepts. Il y en a sans doute d'autres (voir note 3).

phénomène, appelé l'émergence des propriétés (EP) : son explication requiert l'existence de théories et d'exemplaires. Je conclus que les mécanismes qui nous permettent de composer des concepts font appel à différents types de concepts, à savoir des prototypes, des exemplaires et des théories.

## **1. Qu'est-ce que la composition des concepts ?**

### *1.1. La compétence de composition*

#### 1.1.1. L'idée de base

L'affirmation que les concepts se composent revêt des sens différents chez différents auteurs. Dans ce chapitre, je m'intéresse à l'idée suivante. La composition des concepts est la capacité de former des représentations de classes pour lesquelles nous n'avons pas de représentations stockées en MLT. Ces représentations sont formées sur la base des représentations stockées en MLT. En d'autres termes, il y a deux types de concepts, les concepts complexes et les concepts simples. Par définition, un concept est simple si et seulement si c'est une représentation permanente en MLT.<sup>5</sup> Il y a un stock fini de concepts simples. Par exemple, « chien » exprime un concept simple (ou plus probablement plusieurs), puisque j'ai une (ou plusieurs) représentation(s) permanente(s) de chien en MLT. Un concept complexe n'est pas stocké de manière permanente en MLT, mais doit être produit de manière instantanée pour différentes finalités cognitives (par exemple, la compréhension linguistique, le raisonnement contrefactuel...). Dans certaines circonstances, nous devons en effet être capables de produire des concepts que nous ne possédons pas en MLT à partir des concepts que nous possédons. C'est le cas lorsque nous devons catégoriser des objets dans des classes qui sont l'intersection de classes qui sont dénotées par des concepts en MLT, par exemple dans la classe des chiens fous, lorsque nous devons raisonner à propos de ces classes, par exemple faire des raisonnements contrefactuels etc. La thèse que les concepts se composent soutient qu'il existe un mécanisme mental (ou plusieurs) qui opère sur les concepts primitifs afin de produire les concepts complexes. Par exemple, si CHIEN et FOU sont deux concepts

---

<sup>5</sup> Qu'en est-il alors de concepts comme FORFAIT HEBDOMADAIRE qui sont stockés en MLT ? Je soutiens que ce sont des concepts simples. En effet, nous n'avons pas à composer des concepts pour les activer en MT (voir ci-dessous).

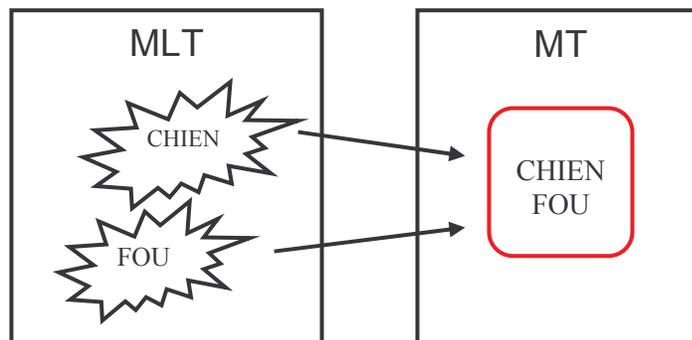
simples, il y a un mécanisme qui prend CHIEN et FOU pour inputs et produit le concept complexe CHIEN FOU. Cette idée est exprimée par le schéma suivant.



**Schéma 6.1.**

Les concepts complexes sont produits en mémoire de travail (MT). Par exemple, « saucisse de chien » exprime un concept complexe pour la plupart des Européens, puisque les Européens ne sont pas habitués à manger du chien et n’ont donc pas de concept de saucisse de chien en MLT. Il en va autrement des Coréens.

J’appelle « catégories complexes » les catégories qui sont dénotées par des concepts complexes. La catégorie des saucisses de chien est une catégorie complexe pour un Européen moyen. Un concept complexe est nouveau pour un sujet si le sujet ne l’a jamais produit auparavant. Une catégorie complexe est nouvelle pour un sujet si elle est dénotée par un concept complexe nouveau. La catégorie des vélos pour dunes est, je suppose, une catégorie complexe nouvelle pour la plupart des lecteurs de ce chapitre.



**Schéma 6.2.**

Comment reconnaît-on un concept complexe ? Comment sait-on qu’une expression exprime un concept complexe, et non un concept stocké en MLT ? Les expressions linguistiques peuvent servir d’indices. Les mots, par exemple, « chien », « chat », expriment habituellement des concepts simples, tandis que de nombreuses

expressions complexes, « saucisses de chien », « vélos pour dunes », expriment des concepts complexes. Cependant, toutes les expressions linguistiques complexes n'expriment pas des concepts complexes. Cela est de toute évidence le cas des idiomes, par exemple « Saint Empire Romain Germanique ». Mais cela est aussi le cas d'expressions non idiomatiques, comme « vin rouge » ou « œuf dur », et peut-être même d'un grand nombre d'expressions linguistiques complexes usuels. Il est peu probable que les concepts qui sont exprimés par de telles expressions soient produits à partir des concepts exprimés par les mots qui composent l'expression complexe, par exemple « vin » et « rouge ». En effet, nous possédons des croyances spécifiques à propos de ces classes, par exemple à propos des vins rouges. Ces croyances sont stockées en MLT.<sup>6</sup> On peut suggérer que plus une expression complexe est fréquente, plus il est probable qu'elle exprime un concept simple. Au contraire, une expression complexe nouvelle, par exemple « grand-mère espionne », exprime un concept complexe.

Si une expression complexe, comme « œuf dur », exprime un concept simple, ce concept n'appartient pas au domaine d'explication d'une théorie de la composition des concepts. En général, les psychologues devraient étudier des concepts exprimés par des expressions comme « grand-mère espionne », plutôt que des concepts exprimés par des expressions comme « vin rouge ». Car il est certain que les concepts exprimés par des expressions complexes nouvelles sont eux-mêmes complexes (Rips 1995).

### 1.1.2. Remarques

Ce chapitre ne s'intéresse pas à l'idée que les concepts sont les briques (« building blocks ») de nos attitudes propositionnelles, c'est-à-dire à l'affirmation que nos concepts sont composés pour produire des croyances, des désirs etc. (voir le chapitre 1, sections 1.2.4. et 1.3.). La notion de composition des concepts réfère dans ce chapitre à la capacité de produire des concepts complexes à partir de concepts simples.

Il existe bien sûr plusieurs types de composition. En particulier, deux concepts peuvent être combinés pour produire un concept complexe (Hampton 1987). « Etudiant d'Oxford qui est maçon » peut exprimer un concept complexe de ce type. J'appelle cette

---

<sup>6</sup> On objecte souvent que l'on sait que les vins rouges sont des vins. Mais cela peut être encodé par un concept de vin rouge.

composition *la combinaison des concepts*. Par ailleurs, un concept peut être modifié en fonction d'un autre concept (Smith et al. 1988). Par exemple, « vélo de plage » peut exprimer un concept de ce type. Le concept qui est exprimé par « plage » n'est pas vraiment combiné avec le concept qui est exprimé par « vélo ». Le concept exprimé par « plage » est utilisé pour modifier le concept exprimé par « vélo ». J'appelle cette composition *la modification des concepts*. Il existe aussi plusieurs types de modification.<sup>7</sup> Par exemple, il faut distinguer les cas où un concept modifie une seule dimension du concept principal, comme dans CHIEN ROUGE, du cas où il en modifie plusieurs, comme dans CHEMISE DE PRINTEMPS. Dans ce qui suit, j'insiste surtout sur la combinaison des concepts.

La composition des concepts ne requiert pas qu'un individu qui possède deux concepts primitifs CHIEN et ROUGE produise le concept complexe CHIEN ROUGE exclusivement à partir de ces deux concepts primitifs. Donc, la composition des concepts n'est pas nécessairement compositionnelle<sup>8</sup> : les propriétés du concept complexe CHIEN ROUGE ne sont pas nécessairement déterminées exclusivement à partir des propriétés des concepts primitifs CHIEN et ROUGE. Cette idée est exprimée par le schéma suivant.



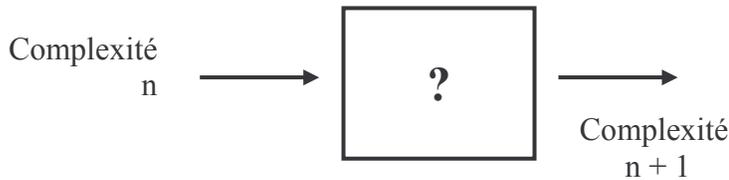
**Schéma 6.3.**

Fodor et d'autres ont soutenu que nous sommes en mesure de produire un nombre infini de concepts complexes : notre compétence de composition est productive (Fodor 1998). Cela est peut-être vrai, mais je ne vais pas insister sur ce point dans ce chapitre.

<sup>7</sup> Machery, E., « Simple heuristics for concept composition », manuscrit.

<sup>8</sup> La notion de compositionnalité peut être utilisée en plusieurs sens. *Stricto sensu*, la compositionnalité est une propriété de la sémantique d'un langage : une sémantique S d'un langage L est compositionnelle si et seulement si pour tout formule non primitive f de L, la signification de f est une fonction de la signification de ces sous-expressions et de la manière dont elles sont composées (pour plus de détails, voir Hodges 2001). Cependant, la notion peut être étendue : un concept complexe est compositionnel si et seulement si ses propriétés sont déterminées par les propriétés des concepts à partir desquels il a été formé (Machery 2002).

Finalement, la compétence de composition est récursive : ses inputs peuvent être des concepts complexes. Elle produit alors des concepts complexes à partir de concepts moins complexes. Par exemple, si « poisson de compagnie » exprime un concept complexe, à savoir POISSON DE COMPAGNIE, ce concept peut servir d'input pour le processus de composition des concepts, par exemple pour produire le concept plus complexe POISSON DE COMPAGNIE DE MAUVAISE HUMEUR.



**Schéma 6.4.**

### 1.2. *Qu'est-ce qu'une théorie de la composition des concepts ?*

Un concept encode des connaissances à propos de son référent. L'objet d'une théorie de la composition des concepts est de découvrir *le ou les mécanismes qui produisent les concepts complexes* et ainsi de déterminer *comment un concept complexe en vient à encoder les connaissances qu'il encode*. Par exemple, elle doit être en mesure d'expliquer comment le concept CHIEN FOU encode l'ensemble des connaissances à propos des chiens fous qu'il encode en fait.

L'objet de ce chapitre n'est pas de faire un examen critique de toute la littérature sur la composition des concepts. Cependant, il importe de souligner une distinction souvent ignorée des psychologues qui travaillent sur la composition des concepts, à savoir la distinction entre la compréhension des expressions complexes et la composition des concepts au sens strict. Le second processus est sans doute une partie du premier : pour comprendre une expression complexe, nous produisons un concept complexe. Mais même si cela est le cas, il faut cependant distinguer le tout, la compréhension d'une expression linguistique complexe, de l'une de ses parties, la production du concept complexe.

Dans les expériences psychologiques sur la composition des concepts, on présente souvent aux sujets des expressions linguistiques complexes nouvelles, par exemple

« grand-mère espionne », « vélos pour dunes » etc. (Kunda et al. 1990, Wisniewski 1996, 2002, Wisniewski et Love 1998, Gagné et Shoben 1997). Les psychologues anglophones ont prêté particulièrement attention aux expressions complexes composées d'un adjectif non intersectif<sup>9</sup> et d'un nom, par exemple « corporate building » ou « open year », et aux expressions complexes formées de deux noms, par exemple « dog magazine », « summer shirt » ou « fork spoon » (Murphy 1988, 1990, Wisniewski 1996, 2002, Wisniewski et Love 1998, Gagné et Shoben 1997, Costello et Keane 2000, 2001, Estes 2003). Car peu de règles linguistiques gouvernent en anglais la compréhension des combinaisons linguistiques adjectif non intersectif-nom et des combinaisons linguistiques nom-nom. Les psychologues ont donc cherché à découvrir les stratégies utilisées par les sujets quand ils interprètent ce type d'expressions complexes. Par exemple, quand on présente à des sujets anglophones l'expression complexe nouvelle « tiger squirrel », comment comprennent-ils cette expression ? Comment associent-ils les deux noms ? Est-ce que les animaux dénotés par « tiger squirrel » sont des écureuils qui *ressemblent* à des tigres ? Ou bien des écureuils qui *se comportent* comme des tigres ? Ou bien quelque chose d'autre ?

Le point important est que la compréhension des expressions complexes inclut la composition des concepts, mais est plus vaste. Par exemple, Murphy soutient que l'interprétation de ce type d'expressions complexes suppose des connaissances théoriques à propos des tigres et des écureuils (Murphy 1988, 1990, 2002, voir aussi Wisniewski 1996). Cela est peut-être vrai : il se peut que des connaissances théoriques soient nécessaires pour interpréter l'expression complexe « dog magazine » comme étant un magazine à propos des chiens (plutôt, par exemple, qu'un magazine lu par des chiens – une interprétation beaucoup moins plausible). Cependant, cela n'est pas un fait à propos du processus de composition des concepts lui-même. C'est un fait à propos de l'interprétation des expressions complexes quand peu de règles linguistiques guident cette interprétation. Expliquer comment les locuteurs comprennent des expressions de ce type consiste entre autres à expliquer comment ils relient les deux noms juxtaposés. Expliquer la production de concepts complexes consiste à expliquer comment il se fait que MAGAZINE A PROPOS DES CHIENS exprime telle et telle connaissance. Les deux

---

<sup>9</sup> Un adjectif A est non intersectif si un AX n'est pas un X qui est un A.

questions sont sans doute intéressantes et peuvent s'éclairer l'une l'autre (Murphy 1988, 1990). Mais les problèmes sont différents.<sup>10</sup>

### 1.3. Ma thèse

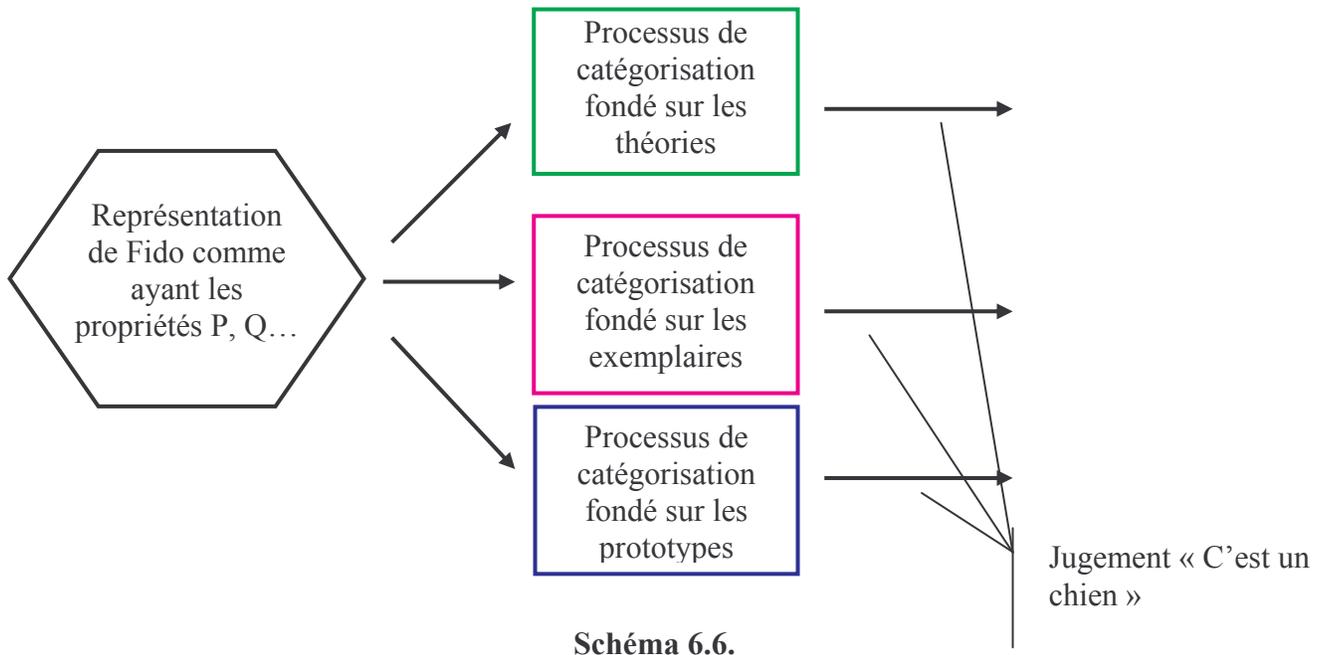
Je soutiens que la compétence de composition repose, entre autres, sur un processus cognitif qui fait appel à des prototypes, à des exemplaires et à des théories. La meilleure explication de certaines compositions des concepts nous conduit donc à postuler l'existence de ces trois types de représentations mentales. Bien sûr, cela vient à l'appui de la prémisse 2 de l'ANE.

Soulignons une différence entre la thèse sur la catégorisation défendue dans le chapitre 5 et la position défendue dans ce chapitre. Dans le chapitre 5, j'ai défendu l'hétérogénéité de la catégorisation : nos performances de catégorisation reposent sur *plusieurs* mécanismes cognitifs, *chacun impliquant un type différent de concepts*, à savoir des prototypes, des exemplaires et des théories. Dans ce chapitre, ma thèse est différente : je soutiens qu'il existe *un* mécanisme de composition des concepts qui implique *à la fois* des prototypes, des exemplaires et des théories.

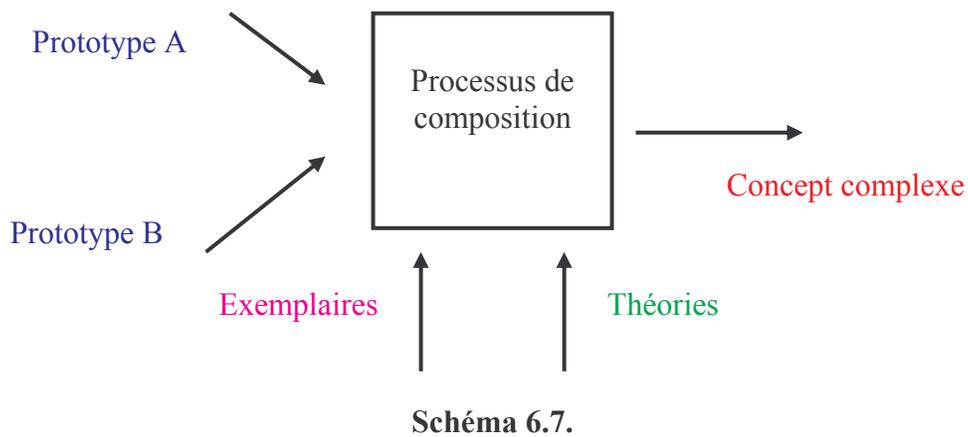
Ainsi, dans le chapitre 5, j'ai affirmé que la catégorisation pouvait être caractérisée par le schéma suivant :

---

<sup>10</sup> La négligence de cette distinction me semble être à l'origine de certains faux débats, par exemple de la critique faite par Gagné et Shoben 1997 de Smith et al. 1988. Le modèle proposé par Smith et al. porte sur la composition des concepts tandis que les données de Gagné et Shoben concernent la compréhension des expressions complexes de la forme nom-nom.



Dans ce chapitre, la position que je défends est plutôt la suivante :



Les deux approches viennent à l'appui de la prémisse 2 de l'ANE et du rejet de l'hypothèse EN : pour expliquer la catégorisation comme la composition des concepts, il faut postuler qu'il existe plusieurs types de concepts, à savoir au moins des prototypes, des exemplaires et des théories.

#### 1.4. Conclusion

Dans cette section, j'ai caractérisé la compétence de composition des concepts que possèdent les êtres humains. Les humains sont capables de produire en mémoire de travail des concepts complexes à partir des concepts qui sont stockés en mémoire à long terme. Comprendre la composition des concepts, c'est expliquer comment il se fait que les concepts complexes encodent les connaissances qu'ils encodent en fait. J'ai finalement présenté la thèse que je souhaite défendre dans ce chapitre : il existe un mécanisme cognitif de composition des concepts qui repose sur les prototypes, les exemplaires et les théories.

## 2. Composition des concepts et prototypes

Dans cette section et dans les sections suivantes, je cherche à montrer que les données empiriques amassées par la recherche psychologique sur la composition des concepts montrent que les inputs du mécanisme de composition des concepts incluent des prototypes, des exemplaires et des théories. Dans cette section, je me penche plus particulièrement sur les prototypes : je montre que les inputs de ce mécanisme incluent des prototypes. En conséquence, il est légitime de postuler des prototypes.

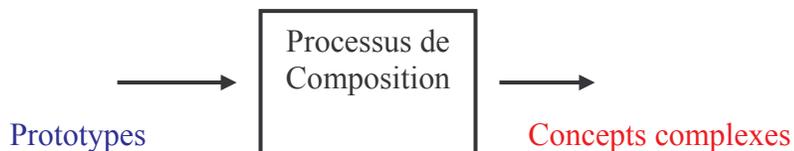


Schéma 6.8.

#### 2.1. Les propriétés typiques des catégories complexes

Quand on présente aux sujets une paire d'expressions linguistiques qu'ils n'ont jamais combinées, par exemple « étudiant d'Oxford » et « maçon », et quand on leur demande de faire la liste des propriétés qu'ils jugent être typiques des étudiants d'Oxford qui sont

des maçons, ils trouvent la tâche naturelle et l'accomplissent souvent sans difficulté (Kunda et al. 1990, Johnson et Keil 2000).<sup>11</sup>

Dans certaines expériences, on a aussi demandé aux sujets de faire la liste des propriétés qui sont typiques des deux catégories originales, par exemple, des étudiants d'Oxford et des maçons (Johnson et Keil 2000 - Hampton 1987 obtient les mêmes résultats avec des concepts complexes qui ne sont sans doute pas nouveaux). Il existe une corrélation entre les propriétés qui sont jugées être typiques des membres des catégories originelles, par exemple des étudiants d'Oxford et des maçons, et les propriétés qui sont jugées typiques des membres de la catégorie complexe, par exemple des étudiants d'Oxford qui sont des maçons (Hampton 1987, Johnson et Keil 2000).

## 2.2. La transmission des propriétés (TP)

Pour expliquer cette observation, nous pouvons proposer le mécanisme suivant (Hampton 1987, 1997b). Nous avons en MLT des représentations qui encodent des connaissances sur les propriétés qui sont jugées être typiques de leurs catégories, à savoir des prototypes. Quand nous produisons un concept complexe, nous extrayons ces prototypes de notre MLT et nous transmettons certaines des propriétés qu'ils représentent au concept complexe. Produire un concept complexe consiste (entre autres, comme nous allons le voir) à déterminer quelles propriétés sont typiques de la catégorie dénotée par le concept complexe, dans notre exemple des étudiants d'Oxford qui sont des maçons, sur la base des prototypes des catégories originelles, dans notre exemple des étudiants d'Oxford et des maçons. Ce processus est connu sous le nom de *transmission des propriétés* (« property inheritance ») :

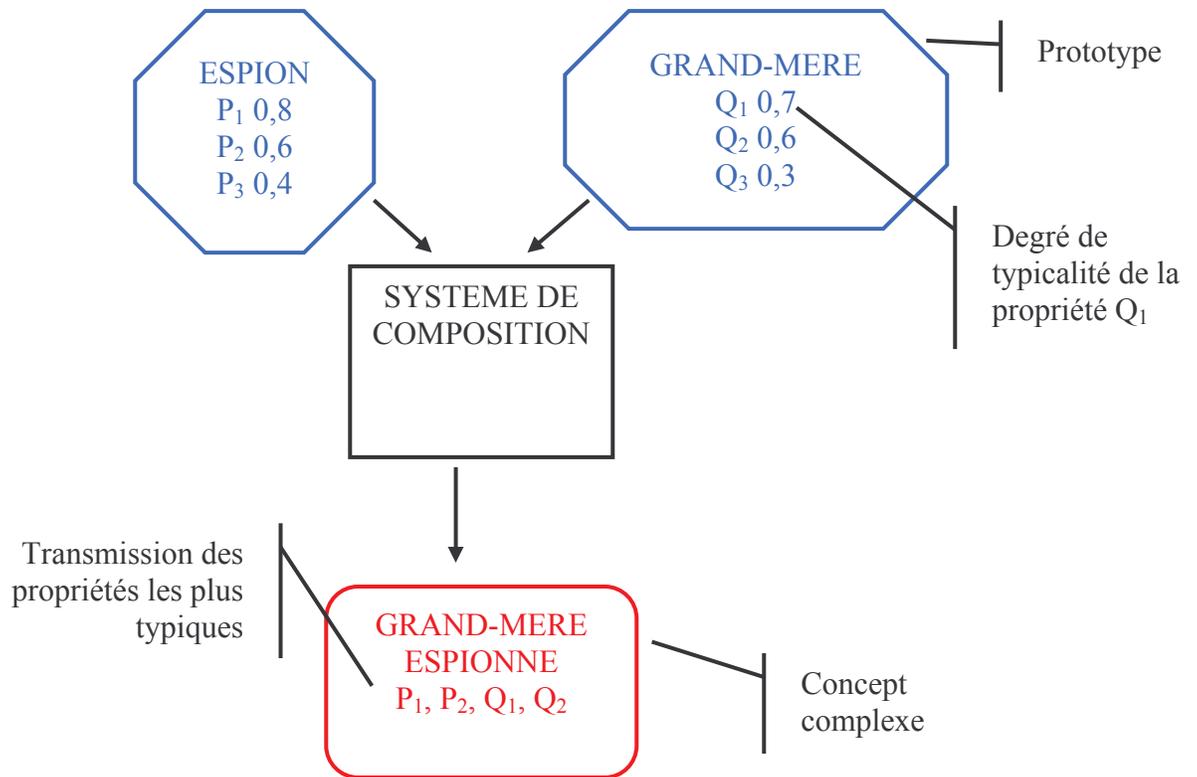
(TP) Parmi les propriétés qui sont représentées par les prototypes originels, certaines sont transmises au concept complexe.

Selon (TP), la transmission des propriétés au concept complexe est déterminée par le principe que les propriétés typiques des membres des catégories dénotées par les

---

<sup>11</sup> Hampton 1987 obtient des résultats semblables avec des concepts complexes qui ne sont sans doute pas nouveaux.

concepts simples sont aussi typiques des membres de la catégorie dénotée par le concept complexe. Cela pose bien sûr la question suivante : *parmi les propriétés typiques, lesquelles sont transmises ?* Comme nous allons le voir dans la section suivante, la question est complexe. Néanmoins, les résultats décrits dans la section 2.1. montrent que le processus de composition suit, entre autres, *la règle de la typicalité élevée* : seules les propriétés qui sont *les plus typiques* des membres des catégories composées sont transmises au concept complexe (Hampton 1987, le principe de diagnosticité, « the diagnosticity assumption », dans Costello et Keane 2000, 2001 exprime une idée semblable). On peut résumer ce mécanisme au moyen du schéma suivant.<sup>12</sup>



**Schéma 6.9. : transmission des propriétés**

<sup>12</sup> Se schéma est *provisoire*. Il ne décrit qu'une partie du processus de composition et ne peut donc rendre compte de tous les phénomènes liés à la composition des concepts.

Les propriétés  $P_3$  et  $Q_3$  ne sont pas transmises, car leur degré de typicalité n'est pas suffisamment élevé. Au contraire, le degré de typicalité de  $P_1$ ,  $P_2$ ,  $Q_1$  et  $Q_2$  est suffisamment élevé. D'après ce mécanisme, les propriétés les plus typiques des grand-mères et les propriétés les plus typiques des espions sont aussi typiques des grand-mères espionnes.<sup>13</sup> Si les grand-mères ont d'habitude les cheveux gris et si les espions maîtrisent les arts martiaux, les grand-mères espionnes ont les cheveux gris et maîtrisent les arts martiaux.

*Soulignons que je ne soutiens pas que les concepts complexes soient des prototypes.* Bien que j'affirme qu'ils encodent des connaissances sur les propriétés qui sont jugées typiques des catégories dénotées, comme le font aussi les prototypes, ils ont des propriétés spécifiques. En particulier, ils encodent des informations modales et causales – comme nous allons le voir ci-dessous. En conséquence, ce ne sont pas des prototypes.

### 2.3. *Objection*

On peut être sceptique. Le fait que les sujets soient capables de faire la liste des propriétés qui sont typiques des étudiants d'Oxford qui sont des maçons, des étudiants d'Oxford et des maçons ne montre pas que leurs *concepts* d'étudiants d'Oxford qui sont des maçons, d'étudiants d'Oxford et de maçons encodent des connaissances à propos de ces propriétés. Après tout, nous avons de nombreuses croyances à propos des classes d'objets dans le monde, par exemple à propos des tables, qui n'appartiennent pas à nos concepts, par exemple au concept TABLE (voir le critère de démarcation, chapitre 1, section 1.2.3.).

Cependant, les propriétés qui sont jugées être typiques des catégories jouent un rôle crucial dans un grand nombre de processus cognitifs. Nous avons déjà vu qu'elles jouent un rôle dans la catégorisation (voir le chapitre précédent, section 4.). Cela est aussi le cas pour les autres processus cognitifs (par exemple, voir Osherson et al. 1990 à propos de l'induction, Sloman 1998 à propos de la déduction, et le chapitre 3, section 2.2.). En d'autres termes, nos connaissances sur ces propriétés constituent des structures de

---

<sup>13</sup> Le fait que cela soit contre intuitif montre que le processus de composition est plus compliqué. Voir la section ci-dessous.

données (« data structures ») qui sont utilisées dans différents processus cognitifs supérieurs. Donc, elles forment des ensembles de connaissances qui, dans certains contextes, sont utilisés par défaut dans nos processus cognitifs supérieurs, bref des concepts (voir le chapitre 1, section 1.2.3.).

On peut cependant se demander si les propriétés qui sont jugées être typiques des catégories dénotées par les concepts complexes, par exemple des étudiants d'Oxford qui sont des maçons, jouent aussi un rôle dans nos processus cognitifs. Quand nous sommes engagés dans un processus cognitif supérieur à propos des étudiants d'Oxford qui sont des maçons, faisons-nous appel aux propriétés que nous jugeons typiques des étudiants d'Oxford qui sont des maçons ? Certains phénomènes empiriques impliquant des concepts complexes sont expliqués de manière satisfaisante, si nous supposons que cela est le cas. Je considère maintenant les données empiriques qui vont dans ce sens.

#### *2.4. Les jugements de typicalité relativement à des catégories complexes nouvelles*

Nous sommes capables d'évaluer la typicalité d'objets par rapport à des catégories complexes nouvelles, c'est-à-dire à des catégories qui sont dénotées par des concepts complexes qui n'ont pas encore été formés, par exemple la catégorie des étudiants d'Oxford qui sont des maçons.

La procédure expérimentale est la suivante (Osherson et Smith 1981, Smith et al. 1988 – Hampton 1988 et Medin et Shoben 1988 utilisent des concepts qui ne sont sans doute pas nouveaux). On présente aux sujets une paire d'expressions qu'ils n'ont probablement jamais combinés. Par exemple, « pomme » et « rayé » pourraient être présentés aux sujets. On présente alors aux sujets des images (ou des descriptions) et les sujets doivent évaluer le degré de typicalité de ces images relativement à la catégorie complexe, par exemple relativement à la catégorie des pommes rayées. Les sujets sont en mesure d'accomplir cette tâche (par exemple, Medin et Shoben 1988).

Or les jugements de typicalité relativement à des catégories simples (c'est-à-dire, représentées par des concepts simples) sont habituellement expliqués en termes de prototypes ou d'exemplaires (voir le chapitre 4, section 2.2. et le chapitre 5, section 2.1.). Puisque la catégorie des pommes rayées est une catégorie complexe nouvelle, ces jugements de typicalité ne reposent pas sur des exemplaires. Il est donc probable qu'ils

reposent sur un concept complexe qui représente les propriétés typiques des pommes rayées. Les concepts complexes représentent donc les propriétés typiques des catégories dénotées.

Cet argument est peut-être cependant un peu rapide. Les jugements de typicalité relativement à des catégories complexes nouvelles ne montrent pas de manière définitive que les concepts complexes encodent des connaissances sur les propriétés typiques des catégories dénotées. Car il se peut que les jugements de typicalité des individus en tant qu'étudiants d'Oxford qui sont des maçons ne reposent pas sur la comparaison de ces individus avec un concept complexe. Ils peuvent dépendre de la typicalité de ces individus en tant qu'étudiants d'Oxford et de leur typicalité en tant que maçons (Huttenlocher et Hedges 1994) : un individu est typique de la catégorie des étudiants d'Oxford qui sont des maçons s'il est un étudiant d'Oxford typique et un maçon typique.

Cette hypothèse pourrait décrire certains jugements de typicalité à propos des instances des concepts complexes. En effet, on a montré que la typicalité des membres de catégories complexes peut parfois être prédite à partir de leur typicalité relativement aux catégories originelles (voir Hampton 1987, 57, Huttenlocher et Hedges 1994). Cela est en particulier le cas quand les sujets croient que l'appartenance à l'une des catégories originelles est indépendante de l'appartenance à l'autre (Huttenlocher et Hedges 1994). Dans ce cas, les sujets admettent que les membres typiques des catégories complexes sont simplement les individus qui sont à la fois typiques des catégories combinées.

Cependant, la typicalité des instances des concepts complexes ne peut d'habitude être ainsi dérivée de leur typicalité par rapport aux catégories originelles. Le processus de composition est tel que les sujets croient souvent que les instances typiques des concepts complexes ne sont pas des instances typiques des concepts combinés. Par exemple, les individus qui sont typiques de la catégorie des étudiants d'Oxford qui sont des maçons ne sont ni des étudiants d'Oxford typiques ni des maçons typiques.<sup>14</sup> En conséquence, les

---

<sup>14</sup> Fodor et Gleitman négligent entièrement ce point dans leurs travaux récents contre l'idée que les prototypes sont impliqués dans la composition des concepts (Connolly, A. C., Fodor, J. A., Gleitman, L. R., et Gleitman, H., 2003, « Why stereotypes don't even make good defaults », manuscrit). Ils croient à tort que cette idée implique que les XY typiques sont des X typiques et des Y typiques.

jugements de typicalité relativement aux concepts complexes ne peuvent être expliqués simplement par référence aux jugements de typicalité relativement aux concepts simples.

Ce phénomène suggère que nous nous représentons bien les propriétés typiques des membres des catégories complexes et que nous utilisons ces propriétés pour produire des jugements de typicalité. Cela vient à l'appui de l'idée que les concepts complexes représentent les propriétés typiques des catégories complexes. Il montre aussi que la règle de la typicalité élevée n'explique pas à elle seule la transmission des propriétés. Autrement, un membre typique de la classe des étudiants d'Oxford qui sont des maçons serait un étudiant d'Oxford typique et un maçon typique. Comme nous allons le voir, la théorie de la composition des concepts proposée dans ce chapitre rend compte de ce phénomène (voir le schéma 6.13.).

### *2.5. Le phénomène de la surextension*

Les données expérimentales résumées ci-dessus pourraient être critiquées comme suit : il est vrai que les sujets sont en mesure de produire des concepts complexes qui représentent des propriétés typiques dans des conditions expérimentales particulières. Mais cela ne montre pas qu'ils produisent habituellement de tels concepts pour représenter les catégories complexes.

Cette objection n'est pas décisive, car l'hypothèse que les sujets produisent des concepts complexes qui représentent les propriétés typiques des catégories dénotées explique aussi d'autres phénomènes psychologiques. Considérons le phénomène de la surextension (« overextension »). La procédure expérimentale est la suivante (Hampton 1982, 1988, 1996, 1997b). On présente aux sujets des énoncés et on leur demande de déterminer s'ils sont vrais ou faux. Par exemple, on présente successivement aux sujets les trois énoncés suivants :

- (1) Les échecs sont un sport qui est un jeu.
- (2) Les échecs sont un jeu.
- (3) Les échecs sont un sport.

Les résultats sont surprenants. Les sujets approuvent les énoncés (1) et (2), mais non l'énoncé (3). Cela est une violation apparente de la logique propositionnelle : si un objet appartient à l'intersection de deux ensembles, il appartient aussi aux deux ensembles. Puisque des objets sont jugés appartenir à l'extension d'un concept complexe de la forme A QUI EST UN B, mais non à l'extension de l'un des concepts composés, par exemple A, Hampton a appelé ce phénomène le phénomène de la surextension.

L'hypothèse selon laquelle nous produisons des concepts complexes qui encodent les propriétés typiques de leurs catégories explique simplement ces résultats. Les énoncés (1) à (3) expriment des jugements de catégorisation : une activité est dite appartenir à une classe d'activités. Si les concepts complexes encodent des connaissances sur les propriétés typiques des catégories dénotées, les jugements de catégorisation reposent probablement, comme dans le PP (voir le chapitre 5, section 2.2.), sur une comparaison entre les objets catégorisés et les propriétés représentées par le concept et sur un calcul de ressemblance. Or, l'énoncé (1) peut être jugé vrai parce que les échecs sont suffisamment semblables au concept complexe des jeux qui sont des sports. L'énoncé (2) peut être jugé vrai parce que les échecs sont suffisamment semblables au prototype des jeux. Cependant, l'énoncé (3) est jugé faux parce que les échecs ne sont pas jugés suffisamment semblables au prototype des sports. Bien sûr, une telle explication suppose l'existence de concepts complexes qui encodent des connaissances sur les propriétés typiques des catégories dénotées.<sup>15</sup>

Aucune de ces expériences n'établit à elle seule que le processus de composition produit, à partir de prototypes, des concepts qui encodent des connaissances sur les propriétés typiques. Cependant, prises ensemble, elles viennent fortement à l'appui de cette idée (pour d'autres phénomènes, voir Hampton 1997b et Murphy 2002).<sup>16</sup>

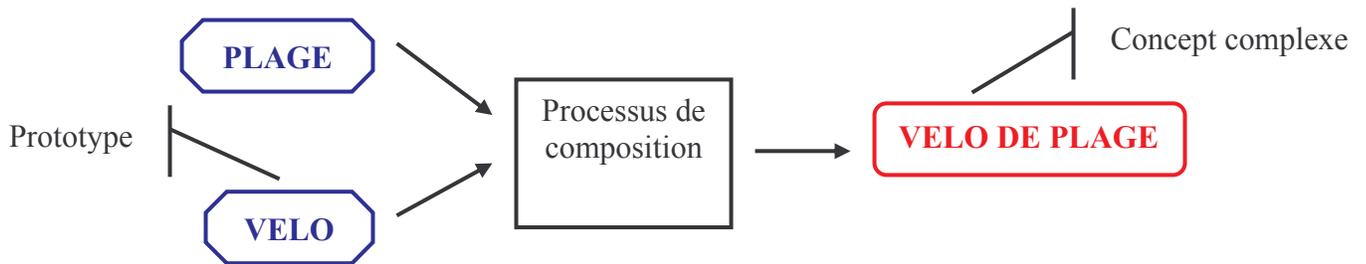
---

<sup>15</sup> Il y a d'autres explications de ce phénomène. Par exemple, la sémantique des propositions relatives peut être particulière. Cette explication prédit que le phénomène de la surextension disparaîtrait si on utilisait la conjonction « et » au lieu d'une relative. Cela doit être testé, même si de prime abord, cela semble douteux. En outre, le phénomène peut être l'effet de règles pragmatiques.

<sup>16</sup> Voir aussi l'erreur de la conjonction décrite dans le chapitre 4, section 2.2.. Si les sujets jugent que Linda, l'étudiante féministe, a plus de chances de devenir une employée de banque féministe qu'une employée de banque, parce qu'ils jugent qu'elle est plus typique de la catégorie des employées de banque féministes que de la catégorie des employées de banque, ils ont probablement produit un concept complexe EMPLOYEE DE BANQUE FEMINISTE qui encode des connaissances sur les propriétés typiques des employées de banque féministes (Smith et Osherson 1989).

## 2.6. Conclusion

Dans cette seconde section, j'ai montré que les inputs du processus de composition des concepts incluent des prototypes. Le schéma suivant résume cette idée :



**Schéma 6.10.**

De nombreuses données montrent que les outputs du processus de composition encodent des connaissances sur les propriétés qui sont jugées être typiques des membres des catégories dénotées : nous utilisons des concepts complexes dans les jugements de typicalité, nous sommes capables de décrire les membres typiques des catégories complexes et certaines propriétés de nos jugements de catégorisation sont bien expliquées par cette hypothèse. De nombreuses propriétés typiques des catégories complexes proviennent des propriétés typiques des catégories combinées. Ce processus est appelé la transmission des propriétés. Il est, entre autres, gouverné par la règle de la typicalité élevée. Ces données montrent que les inputs du processus de composition incluent des prototypes. Il est donc légitime de postuler des prototypes.

### 3. Les théories et la transmission des propriétés

La description du processus de composition proposée ci-dessus est incomplète. En effet, les instances des concepts complexes sont parfois jugées être des membres atypiques des catégories combinées. Si la TP était seulement gouvernée par la règle de la typicalité élevée, les instances typiques des concepts complexes seraient toujours des membres typiques des catégories combinées. Cela suggère donc que la TP n'est pas seulement gouvernée par la typicalité des propriétés. D'autres considérations interviennent pour déterminer quelles propriétés sont représentées par les concepts complexes. Comme nous allons le voir, la TP repose aussi sur des concepts théoriques.

### 3.1. La TP et les connaissances modales

Quelles propriétés sont transmises aux concepts complexes ? Des propriétés qui sont jugées être typiques des classes composées, comme le montrent les recherches de Hampton. Mais ce n'est pas le seul type de propriétés : nous utilisons aussi *nos connaissances modales à propos des catégories originelles* pour sélectionner les propriétés *nécessaires* et pour exclure les propriétés *impossibles*.

Considérons d'abord *les propriétés nécessaires* (Hampton 1987). Hampton a montré que quand nous croyons qu'une propriété P est nécessairement possédée par les membres d'une catégorie C dénotée par un ou plusieurs concepts, nous transmettons toujours P aux concepts complexes produits à partir des concepts de C. Par exemple, si nous croyons que les grand-mères sont nécessairement des mères de parents, la propriété *être une mère de parent* est toujours transmise aux concepts complexes produits à partir des concepts de grand-mères, par exemple au concept de grand-mère espionne.

Considérons maintenant *les propriétés impossibles* (Hampton 1987). Les propriétés qui sont jugées ne pouvoir être possédées par les membres d'une catégorie C ne sont jamais représentées par les concepts complexes produits à partir d'un concept de C. Par exemple, *migre* est une propriété typique des oiseaux. Cependant, elle n'est pas transmise au concept complexe OISEAU EN CAGE. L'explication est la suivante : les animaux qui vivent en cage ne peuvent migrer.

En résumé, quand deux prototypes sont combinés ou quand un prototype est modifié par un autre concept, les propriétés qui sont jugées nécessaires sont toujours transmises et les propriétés qui sont jugées impossibles ne sont jamais transmises. Or, les prototypes n'encodent pas des connaissances sur les propriétés nécessaires ou impossibles des membres des classes dénotées, mais seulement sur leur fréquence. En d'autres termes, les prototypes n'encodent pas de connaissances modales, mais seulement des connaissances statistiques. En conséquence, des connaissances qui ne sont pas stockées dans les prototypes sont utilisées lors de la composition des concepts. Or, les concepts théoriques sont précisément supposés encoder des connaissances modales. Cela suggère que quand un prototype des A et un prototype des B sont combinés, des concepts théoriques des A et des B sont aussi impliqués pour produire le concept complexe : ils

sélectionnent les propriétés qui sont nécessaires et excluent les propriétés qui sont impossibles.

### 3.2. La TP et les connaissances causales

Quand nous produisons un concept complexe, nous distinguons aussi *les propriétés qui sont causalement centrales* (Medin et Shoben 1988, Murphy 1988, 1990, Johnson et Keil 2000). Une propriété est causalement centrale si elle est causalement ou fonctionnellement reliée aux autres propriétés possédées par les membres de la catégorie (voir le chapitre 5, section 3.2.). Par exemple, *être incurvé* est une propriété causalement centrale de la catégorie des boomerangs. Car les sujets croient (à tort) qu'un boomerang peut revenir à son lanceur, c'est-à-dire peut satisfaire sa fonction, *parce qu'il est incurvé*. Au contraire, *être fait de bois* n'est pas une propriété causalement centrale : les sujets croient que les boomerangs pourraient être faits d'une autre matière, bien que les boomerangs soient en général faits de bois (Medin et Shoben 1988, 175-176).

Or les propriétés qui sont causalement centrales ont plus de chances d'être représentées par les concepts complexes. Par exemple, la propriété *être incurvé* a plus de chances que la propriété *être fait de bois* d'être représentée par un concept complexe fait à partir d'un concept de boomerang, même si la typicalité des deux propriétés est la même (Medin et Shoben 1988).

L'argument est le même que précédemment. Les prototypes n'encodent pas des connaissances causales sur les propriétés des membres des catégories dénotées. En conséquence, la composition des concepts repose sur un type de connaissances qui n'est pas encodé par les prototypes. Les concepts théoriques sont supposés encoder ce type de connaissances. Cela suggère que quand un prototype des A et un prototype des B sont composés, les concepts théoriques des A et des B sont impliqués dans la production du concept complexe : ils distinguent les propriétés qui sont causalement centrales de celles qui ne le sont pas.

En conséquence, la typicalité n'est pas le seul facteur qui détermine quelles propriétés sont représentées par les concepts complexes. Nous prenons bien en considération la typicalité des propriétés (voir la règle de la typicalité élevée). Mais nous

prenons aussi en compte d'autres facteurs. Nous distinguons les propriétés qui sont nécessaires, impossibles et causalement centrales. Or, cela requiert un type de connaissances qui n'est pas encodé par les prototypes. Les concepts théoriques sont supposés encoder ces connaissances. Cela suggère que le processus de composition repose aussi sur des concepts théoriques des catégories qui sont composées.

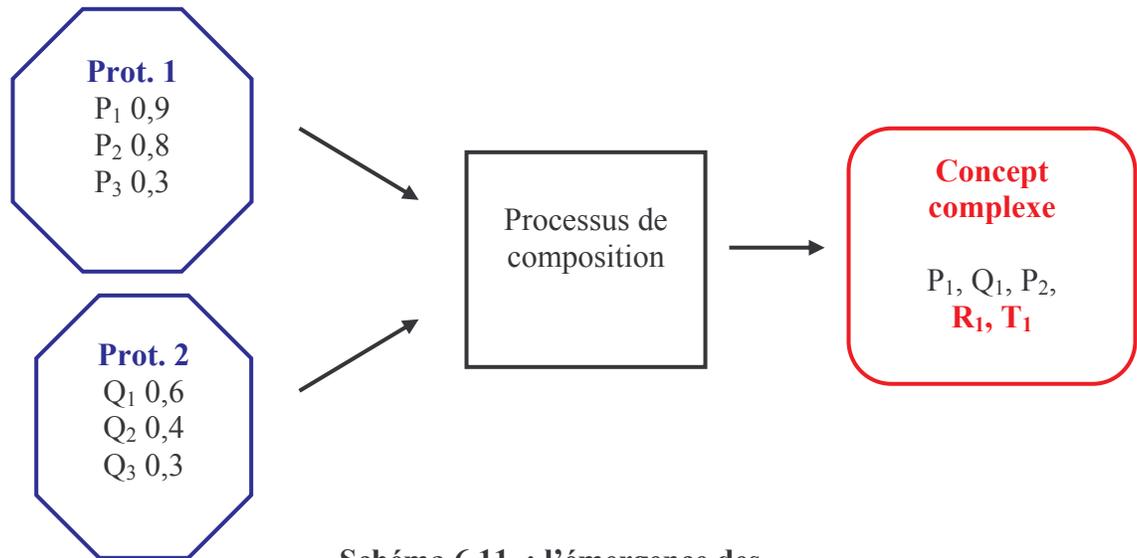
### 3.3. Conclusion

La composition des concepts repose sur des connaissances modales et causales : certaines propriétés sont distinguées, parce qu'elles sont jugées impossibles, nécessaires ou causalement centrales. Les prototypes n'encodent pas ce type de connaissances. Cela suggère que la composition des concepts prend pour inputs des concepts théoriques.

## 4. L'émergence des propriétés (E<sub>m</sub>P)

### 4.1. Qu'est-ce que l'émergence des propriétés ?

Quand nous produisons un concept complexe à partir de deux prototypes, certaines connaissances qui sont encodées par le concept complexe ne sont pas héritées des prototypes composés (Hampton 1987, 1997b, Kunda et al. 1990, Johnson et Keil 2000). L'émergence des propriétés est donc le phénomène suivant. Quand on demande aux sujets de faire la liste des propriétés qu'ils supposent être typiques d'une catégorie complexe, *toutes les propriétés qui sont proposées pour la catégorie complexe (par exemple, les étudiants d'Oxford qui sont des maçons) ne sont pas tirées des propriétés qui sont proposées pour les catégories composées (les étudiants d'Oxford et les maçons)*. Hampton a nommé ces propriétés « les propriétés émergentes ». Par exemple, la propriété *généreux* est jugée être typique des étudiants d'Oxford qui sont des maçons, bien que ni les étudiants d'Oxford ni les maçons ne soient jugés être généreux. Le schéma ci-dessous résume ce phénomène (R<sub>1</sub> et T<sub>1</sub> sont les propriétés émergentes).



**Schéma 6.11. : l'émergence des propriétés**

D'où viennent ces propriétés ? Il y a deux explications de l'émergence des propriétés. Elles ne sont pas incompatibles.

#### *4.2. Le feedback extensionnel : propriétés émergentes et exemplaires*

##### 4.2.1. Le feedback extensionnel

Le premier type de propriétés émergentes est tiré de représentations d'individus singuliers. Hampton a appelé ce phénomène le feedback extensionnel (Hampton 1988). Par exemple, il est probable que l'expression « étudiant de l'ENS qui est un O.S. » n'exprime pas un concept simple. Un concept complexe doit donc être produit à partir (entre autres) des prototypes des étudiants de l'ENS et des O.S.. Cependant, si je devais faire la liste des propriétés des étudiants de l'ENS qui sont des O.S., je mentionnerais des propriétés que je ne mentionnerais ni pour les étudiants de l'ENS ni pour les O.S.. L'explication est sans doute la suivante : je connais un étudiant de l'ENS qui est un O.S., à savoir Robert Linhart (voir *L'Etabli*). J'ai donc en MLT une représentation singulière de Linhart. Quand je construis le concept complexe, j'extrais de ma MLT cette représentation singulière et j'encode dans le concept complexe les propriétés représentées

par la représentation singulière de Linhart, par exemple l'engagement à l'extrême gauche.  
Le schéma suivant résume cette idée.

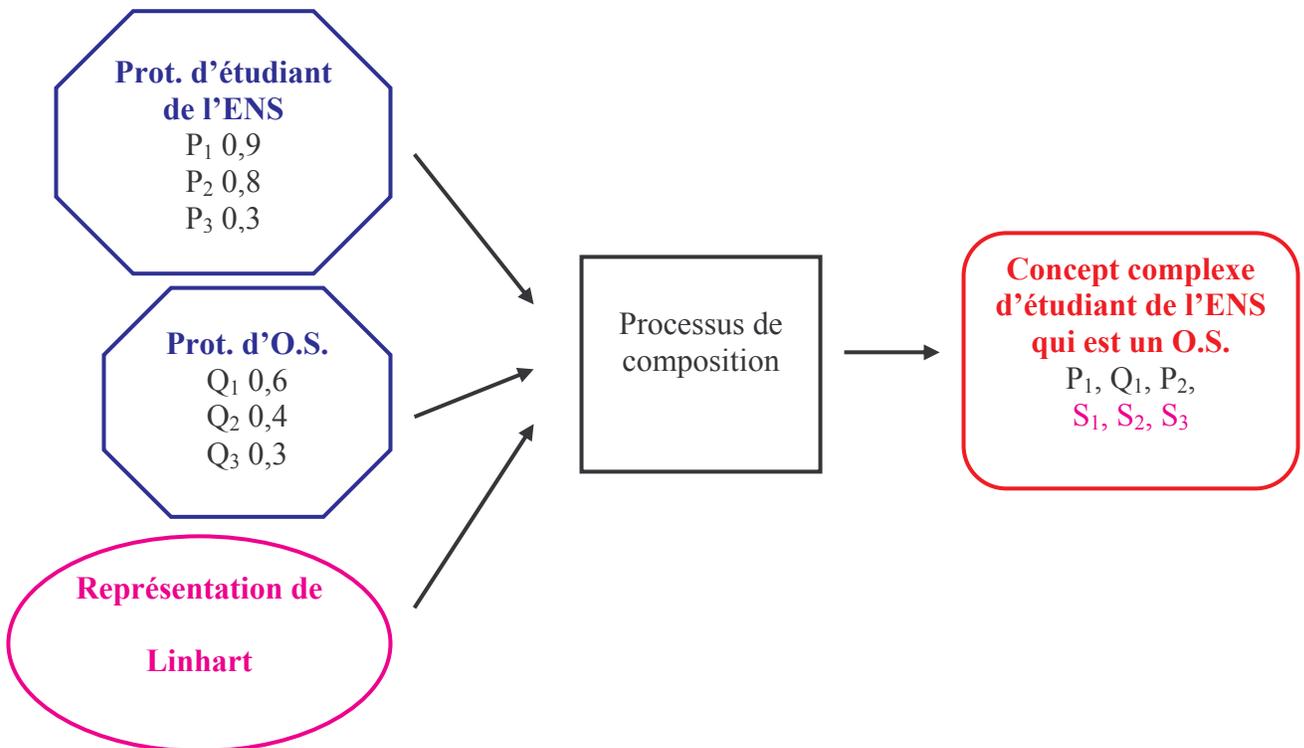


Schéma 6.12.

#### 4.2.2. Feedback extensionnel et exemplaires

Dans notre exemple, la représentation singulière de Linhart sert d'exemplaire de la catégorie des étudiants de l'ENS qui sont des O.S. : les autres membres de la catégorie sont supposés ressembler à Linhart.

#### 4.2.3. Problèmes

On pourrait avancer plusieurs objections contre cette proposition. Examinons-les. Tout d'abord, comment nos représentations singulières sont-elles extraites ? Si je dois chercher en MLT une instance du concept complexe ETUDIANT DE L'ENS QUI EST UN O.S., la proposition semble circulaire : je devrais déjà avoir formé un concept complexe afin de déterminer ses instances. Mais cela n'est pas nécessairement le cas. Quand je combine deux prototypes, A et B, il se peut que je cherche des représentations singulières qui encodent la proposition que l'individu est une instance de A (par exemple, la proposition

[Linhart est un étudiant de l'ENS]) et la proposition que l'individu est une instance de B (par exemple, la proposition [Linhart est un O.S.]). Autrement dit, je peux chercher en MLT des instances des deux concepts. En conséquence, pour extraire les instances d'un concept complexe, il n'est pas nécessaire que j'aie formé le concept complexe lui-même. La proposition n'est donc pas circulaire.

Il y a d'autres problèmes cependant. En particulier, si je possède une représentation singulière qui peut servir de représentant pour la catégorie complexe, pourquoi produire un concept complexe à partir de cet exemplaire ? Pourquoi ne pas l'utiliser directement ? Développons ce problème. Si j'enrichis le concept complexe au moyen des propriétés représentées par la représentation singulière, par exemple la représentation de Linhart, je suppose par là que les membres de la catégorie complexe ressemblent à l'individu représenté. Mais si je fais cette hypothèse, pourquoi ne pas utiliser la représentation singulière dans mes processus cognitifs, c'est-à-dire dans mes raisonnements, inférences etc., plutôt que de produire un concept complexe à partir d'elle ? Si je n'utilise pas cette représentation singulière dans mes processus cognitifs, c'est probablement que je suppose que les membres de la catégorie complexe ne ressemblent pas à l'individu représenté. Mais alors, pourquoi prends-je en considération cette représentation singulière pour enrichir le concept complexe ?

Je soutiens que nous représentons souvent une catégorie complexe au moyen d'exemplaires. De ce fait, il est probable qu'à l'occasion nous utilisons directement une représentation singulière dans nos processus cognitifs à propos des catégories complexes (voir Prinz 2002, Machery<sup>17</sup>). Cependant, parfois, il peut être rationnel de produire un concept complexe à partir de prototypes et de théories en l'enrichissant des propriétés représentées par une représentation singulière plutôt que d'utiliser directement cette représentation dans nos processus cognitifs. Par exemple, supposons que je n'ai en MLT *qu'une seule* représentation singulière d'un membre de la catégorie complexe. Plutôt que d'assimiler les membres de la catégorie complexe à cette unique représentation, il peut être plus rationnel de prendre aussi en compte les prototypes et les concepts théoriques des catégories combinées pour former le concept complexe – en bref, de supposer que les

---

<sup>17</sup> Machery, E., « Simple heuristics for concept composition », manuscrit.

membres de la catégorie complexe possède les propriétés typiques, nécessaires et causalement centrales des catégories combinées *et* ressemblent à cet individu singulier. Car cela augmente la probabilité que ma représentation de la catégorie complexe soit correcte.

#### 4.3. Propriétés émergentes et théories

D'autres propriétés émergentes ne peuvent être ainsi expliquées. Car elles apparaissent dans des concepts complexes dont nous ne connaissons aucune instance. Ainsi, la plupart des sujets n'ont jamais rencontré d'étudiant d'Oxford qui sont des maçons. Cependant, ils soutiennent que ces individus ne sont pas attirés par l'argent, sont généreux ou ont un tempérament artistique (Kunda et al. 1990).

Ces propriétés émergentes sont le résultat d'un raisonnement causal à propos des catégories qui sont combinées (Kunda et al. 1990, Johnson et Keil 2000). Ce raisonnement causal implique des concepts théoriques des catégories composées. Un exemple peut éclairer cette idée. Pourquoi le concept des étudiants d'Oxford qui sont des maçons représente-t-il la propriété *ne pas être attiré par l'argent* (Kunda et al. 1990) ? Kunda et al. (1990) ont interviewé des sujets. Ces entretiens suggèrent que ces sujets raisonnent causalement à propos de l'appartenance aux catégories complexes. On peut reconstruire leur raisonnement causal comme suit. Le prototype ETUDIANT D'OXFORD représente certaines propriétés qui sont jugées improbables pour un maçon : par exemple, le sujets jugent que les étudiants d'Oxford sont ambitieux. Les sujets cherchent donc à expliquer comment un individu peut être à la fois un étudiant d'Oxford et un maçon. Certains concluent donc que cet individu n'est probablement pas attiré par l'argent etc. Il en résulte que le concept complexe représente la propriété *ne pas être attiré par l'argent*.<sup>18</sup>

Un tel processus suppose l'existence de concepts théoriques des étudiants d'Oxford et des maçons. Car les prototypes ne figurent pas dans des raisonnements causaux, puisqu'ils n'encodent que des connaissances statistiques. Il est donc plausible que des concepts théoriques servent d'inputs au processus de composition.

---

<sup>18</sup> Les expériences de Johnson et Keil 2000 confirment cette interprétation des entretiens.

#### *4.4. Conclusion*

Dans cette section, j'ai décrit un phénomène caractéristique de la production de concepts complexes, appelé l'émergence des propriétés. Certaines propriétés sont jugées être typiques d'une catégorie complexe, alors qu'elles ne sont pas jugées être typiques des catégories combinées. J'ai fourni deux explications de ce phénomène. La première repose sur des représentations singulières qui servent d'exemplaires de la catégorie complexe, la seconde sur des concepts théoriques. En conséquence, pour expliquer la composition des concepts, nous devons supposer l'existence d'exemplaires et de théories.

### **5. Un modèle de la composition des concepts**

#### *5.1. Le modèle*

Je propose ci-dessous un modèle du processus de composition des concepts qui intègre les éléments d'explication avancés dans les sections précédentes. Je n'ai pas l'intention d'expliquer par là tous les phénomènes qui caractérisent la composition des concepts. Bien du travail reste à faire (c'est aussi la conclusion de Murphy 2002, chapitre 12). Cependant, il est possible que le schéma 6.13. décrive certains traits essentiels du processus de composition des concepts.

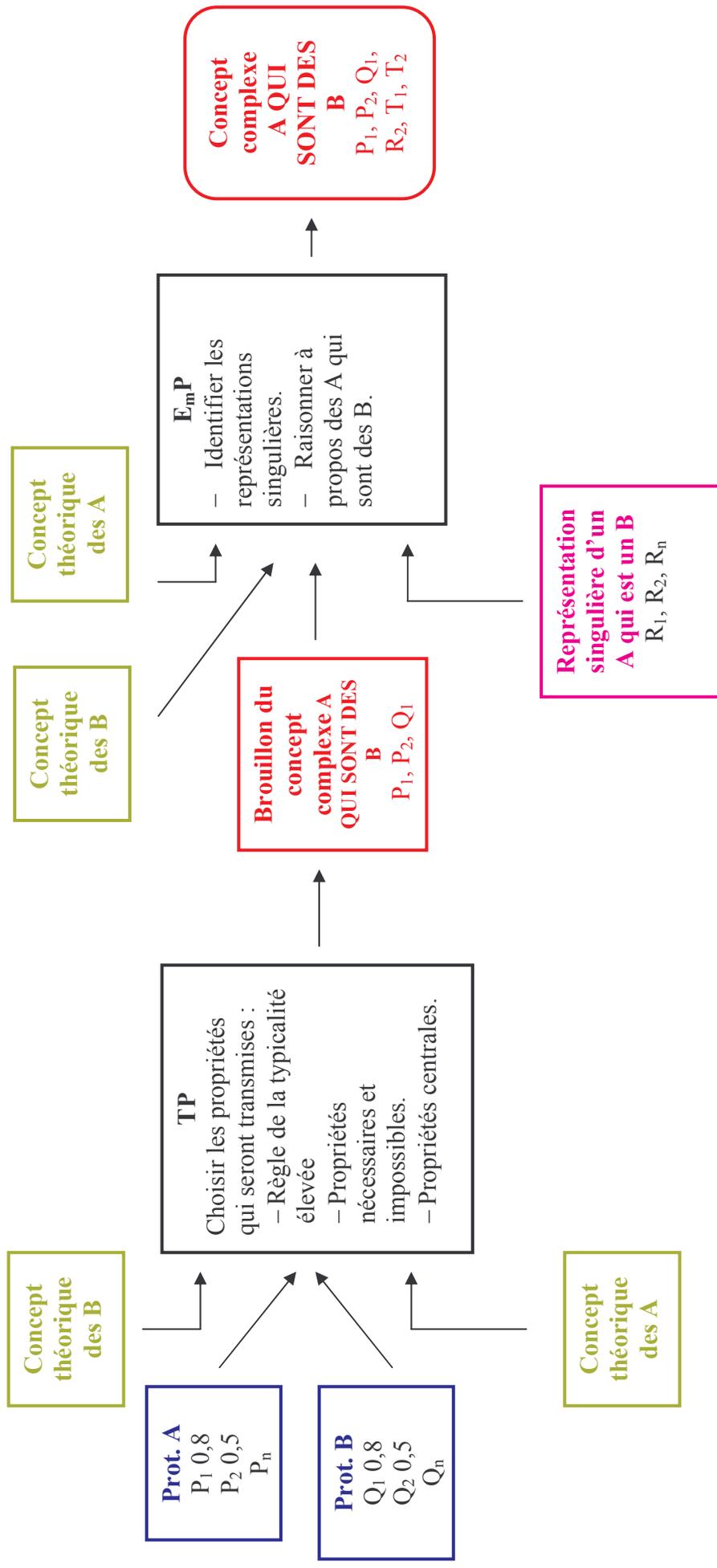


Schéma 6.13. :  
Résumé du processus de composition

## 5.2. Commentaires

Un concept complexe est produit en deux étapes. Dans un premier temps, les propriétés représentées par les deux prototypes combinés sont transmises à un brouillon du concept complexe. Cette transmission obéit aux règles suivantes :

- La règle de la typicalité élevée : les propriétés dont la typicalité est élevée sont transmises.
- Les propriétés qui sont jugées nécessaires pour l'une des catégories combinées sont toujours transmises.
- Les propriétés qui sont jugées impossibles pour l'une des catégories combinées ne sont jamais transmises.
- Les propriétés qui sont jugées être causalement centrales pour l'une des catégories combinées sont transmises de manière préférentielle.

Les trois dernières règles requièrent des concepts théoriques des catégories qui sont combinées. La première règle requiert des prototypes. Le modèle composite de Hampton 1987 (« the composite model »), le modèle de la spécialisation des concepts de Murphy 1990 (« the concept specialization model ») et le modèle C<sup>3</sup> de Costello et Keane 2000 décrivent certains aspects de cette étape (voir aussi Prinz 2002, chapitre 12).

Dans la seconde étape, le modèle est enrichi de propriétés émergentes. Deux processus sont pertinents :

- Feedback extensionnel : des représentations singulières sont extraites de la MLT.
- Raisonnement causal : nous raisonnons causalement à propos des catégories complexes.

## 6. Conclusion : la composition des concepts et la prémisse 2 de l'ANE

L'hypothèse que nous utilisons plusieurs types de représentations mentales dans nos processus cognitifs supérieurs est présupposée par le modèle de la composition des concepts résumé ci-dessus (schéma 6.13.). Puisque ce modèle offre une explication convaincante des données expérimentales à propos de la composition des concepts, l'étude empirique de la composition des concepts vient donc à l'appui de l'hypothèse de l'hétérogénéité. Cependant,

cette conclusion n'est justifiée que si le modèle présenté ci-dessus est la meilleure explication des données empiriques. Est-ce le cas ?

De toute évidence, les données empiriques à propos de la composition des concepts ne pourraient être expliquées par une théorie qui ne poserait que des exemplaires, que des prototypes ou que des théories. Cependant, pour les raisons expliquées à la fin du chapitre précédent (section 5.), certaines théories composites des concepts peuvent expliquer aussi bien les données empiriques sur la composition des concepts que l'hypothèse de l'hétérogénéité. Il serait en effet assez simple de reformuler le schéma 6.13. en s'appuyant sur une théorie composite des concepts : il suffit seulement de supposer que les concepts ont trois parties, l'une encodant des connaissances statistiques, la seconde encodant des connaissances à propos des individus et la troisième encodant des connaissances théoriques.

J'examinerai cette objection en détail dans le chapitre 8. La conclusion ne peut donc être que provisoire : à l'exception de certaines théories composites des concepts, le modèle présenté dans la section 5., qui repose sur l'hypothèse de l'hétérogénéité, est la meilleure explication des données empiriques à propos de la composition des concepts. La prémisse 2 de l'ANE est donc conditionnellement établie.

## Chapitre 7

### Est-ce que l'eau est H<sub>2</sub>O ?

---

Dans les chapitres 5 et 6, j'ai cherché à défendre la prémisse 2 de l'argument de la nécessité explicative (ANE) : il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs. L'étude empirique de deux processus cognitifs, la catégorisation (chapitre 5) et la composition des concepts (chapitre 6) montrent que des prototypes, des exemplaires et des théories sont utilisés dans nos processus cognitifs supérieurs. Il aurait été possible de me tourner maintenant vers une autre compétence cognitive supérieure, par exemple vers l'induction ou la capacité de faire des analogies. J'aurais cherché à montrer que la meilleure explication des propriétés de ces compétences postule l'existence de ces trois types de représentations mentales.

Au lieu de cela, je présente dans ce chapitre un autre type de données. Supposons que l'usage des expressions linguistiques (ou au moins de certains types d'expressions, par exemple les noms communs ou les verbes) soit influencé par les concepts auxquels elles sont associées.<sup>1</sup> L'hypothèse de l'hétérogénéité prédit alors que de nombreuses expressions linguistiques, par exemple les noms communs, sont utilisées de manière hétérogène. Car l'hypothèse de l'hétérogénéité soutient que de nombreuses catégories, par exemple la catégorie des chiens, sont dénotées par plusieurs concepts qui appartiennent à des types de représentations mentales qui ont peu en commun (voir le chapitre 3). En conséquence, « chien » peut être associé à différents concepts, en particulier à un prototype de chiens, CHIEN<sub>Pr</sub>, à un ensemble d'exemplaires de chiens, CHIEN<sub>Ex</sub>, et à une théorie à propos des chiens, CHIEN<sub>T</sub>. Dans certaines circonstances, « chien » peut être associé à CHIEN<sub>Pr</sub>, dans d'autres circonstances, à CHIEN<sub>Ex</sub>, et dans d'autres circonstances encore, à CHIEN<sub>T</sub>. Comme ces concepts sont différents, le nom « chien » est sans doute utilisé différemment, quand il est associé à ces différents concepts. Dans ce chapitre, je défends cette prédiction.

Je me concentre sur un nom commun particulier, « eau ». Je montre d'abord que le comportement linguistique des locuteurs est paradoxal : alors qu'ils soutiennent que l'eau est par essence H<sub>2</sub>O, ils utilisent « eau » pour nommer des liquides sans tenir compte de leur

---

<sup>1</sup> Fodor 1998 soutient que la signification d'une expression linguistique est le concept qu'elle exprime (mais voir la critique de Prinz 2002). Nombre de psychologues soutiennent plutôt que la signification et l'usage des expressions linguistiques sont influencés par les concepts associés (Clark 1983, voir aussi Keil 1989, chap. 7, 1991, Chomsky 1995). Je ne me prononce pas sur ce point.

composition chimique. Leur comportement linguistique est donc dissocié de leur croyance explicite. Bien que cette dissociation ait été interprétée de plusieurs manières dans la littérature philosophique et psychologique récente, je montre qu'elle vient en fait à l'appui de la thèse selon laquelle « eau » est associé au moins à deux concepts différents et selon laquelle l'un est une théorie et l'autre un prototype ou un ensemble d'exemplaires. Bien que cette dissociation ne permette pas d'établir l'existence de *trois* représentations mentales, à savoir d'un prototype d'eau, d'exemplaires d'eaux et d'une théorie à propos de l'eau, elle est compatible avec la prédiction dérivée ci-dessus à partir de l'hypothèse de l'hétérogénéité.

Dans la première section de ce chapitre, j'établis l'existence d'une dissociation entre notre croyance sur l'essence de l'eau et certains usages du mot « eau ». Dans la seconde section, je critique certaines interprétations récentes de ce phénomène. Dans la troisième section, je montre que cette dissociation vient à l'appui de la prémisse 2 de l'ANE.

## 1. Eau, sang et thé

### 1.1. L'essentialisme psychologique (EP)

L'essentialisme psychologique est un phénomène psychologique qui a été décrit pour la première fois par Medin et Ortony (1989). Il est rapidement devenu l'un des phénomènes psychologiques les mieux étudiés en psychologie des concepts (Medin 1989, Atran 1990, Gelman et Coley 1991, Gelman et Wellman 1991, Gelman et al. 1994, Gopnik et Meltzoff 1997, chap. 6, Keil et Richardson 1999, voir particulièrement la recension de Gelman et Hirschfeld 1999 et la discussion de Strevens 2000, 2001). L'EP a deux aspects :

- Tout d'abord, l'essentialisme psychologique est *une croyance sur l'identité des catégories*. L'idée principale est que les individus croient que la possession de certaines propriétés est nécessaire et suffisante pour appartenir à certaines catégories. Être essentialiste à propos des oiseaux, par exemple, c'est croire qu'il y a certaines propriétés qui déterminent ce que c'est que d'être un oiseau : pour être un oiseau, il est nécessaire et suffisant de posséder ces propriétés.<sup>2</sup>

Par exemple, Gelman et Wellman décrivent cet aspect de l'EP comme suit (1991/1999, 624) :

---

<sup>2</sup> Quelle est la différence entre le paradigme classique et la théorie de l'essentialisme psychologique ? Selon le PC, nos concepts spécifient des conditions nécessaires et suffisantes d'appartenance. Selon la théorie de l'EP, nos concepts incluent la croyance qu'il existe des conditions nécessaires et suffisantes sans nécessairement spécifier ces conditions. Voir ci-dessous.

« An essence is a unique, typically hidden property of an object that makes it what it is, without which it would have a different identity (e.g., the chemical composition of water, the DNA structure of an elephant). »

- Deuxièmement, l'essentialisme psychologique est une croyance sur la cause des propriétés qui sont typiques des membres d'une catégorie (comportements typiques, apparence typique...). Etre essentialiste à propos des oiseaux, c'est croire que de nombreux oiseaux possèdent les propriétés caractéristiques des oiseaux, parce leur essence cause la possession de ces propriétés. Keil et Richardson décrivent, par exemple, cet aspect de l'EP comme suit (1999, 271) :

« *causal essence* refers to something about a kind that results in its having many of its most typical and stable properties. »

Bloom résume habilement ces deux aspects comme suit (ms) :

« Psychological essentialism is the theory that people tacitly believe that certain categories have 'essences' that determine category membership and are, in the normal course of affairs, causally responsible for observable properties of category members. For instance, people might believe that all tigers share a hidden property that causes them to look and act like tigers. »<sup>3</sup>

La nature de l'essence d'une catégorie n'est pas nécessairement connue. Nous pouvons croire que les membres d'une catégorie possèdent une essence, sans avoir la moindre idée à propos de la nature de cette essence. Medin et Ortony (1989, 184-185) ont utilisé la notion de garde-place (« placeholder ») pour décrire ce phénomène : un concept essentialiste de chiens que possède un individu contient un garde-place si cet individu croit que les chiens ont une essence, tout en reconnaissant ignorer la nature de cette essence. Dans d'autres cas, nous pouvons avoir des croyances vagues sur la nature de cette essence (Keil et Richardson

---

<sup>3</sup> Locke a élaboré philosophiquement cette croyance populaire. Il écrit en effet (1690/1964, 286) : « By this real essence, I mean that real constitution of anything, which is the foundation of all those properties that are combined in, and are constantly found to coexist with, the nominal essence: that particular construction which everything has within itself, without any relation to anything about it. » Par « foundation », Locke veut dire « cause ».

1999).<sup>4</sup> Par exemple, les jeunes enfants semblent croire que l'essence des animaux est une partie physique interne (Gelman et Wellman 1991). Les adultes peuvent croire que l'essence des espèces est codée par leurs gènes – sans, bien sûr, être capable d'expliquer cette croyance erronée. Finalement, nous pouvons avoir des croyances précises sur la nature des essences de certaines catégories : par exemple, nous croyons que l'essence de l'eau, c'est d'être composée de molécules d'H<sub>2</sub>O (voir ci-dessous).

Par ailleurs, l'EP peut porter sur tous les domaines (« domain-general ») ou porter sur certains domaines (« domain-specific », voir Atran 1990 et la discussion dans Hirschfeld et Gelman 1999). Si l'EP porte sur tous les domaines, tous nos concepts, dans n'importe quel domaine, incluent des croyances essentialistes, par exemple les concepts d'artéfacts (TABLE...), les concepts de matières (OR...) ou les concepts d'animaux (LION...). Au contraire, si l'EP porte sur certains domaines particuliers, par exemple sur le domaine des animaux, seuls certains concepts incluent des croyances essentialistes, par exemple CHIEN ou RAT, mais pas TABLE ou LUSTRE (par exemple, Keil 1989, Atran 1990). Les deux positions admettent en fait que nous avons des concepts essentialistes à propos du domaine des animaux, du domaine des plantes et du domaine des matières. C'est à propos d'autres domaines, par exemple du domaine des artéfacts ou du domaine des groupes sociaux (Hirschfeld 1996, Gil-White 1999, 2001, Machery et Faucher<sup>5</sup>), qu'elles divergent. Dans ce chapitre, je prête particulièrement attention à une matière, l'eau. Je peux donc rester neutre sur cette question.

Finalement, quel est le rapport entre l'EP et les types de concepts que j'ai distingués dans le présent travail (prototypes, ensembles d'exemplaires, concepts théoriques) ? Comme nous le verrons plus loin (voir aussi le chapitre 3, section 2.1.), admettre des concepts essentialistes, ce n'est pas admettre un nouveau type de concepts, fort différent des autres types de concepts (voir la section 3.4.). Les concepts essentialistes sont en effet des concepts théoriques.

De nombreuses données empiriques montrent que nous sommes essentialistes (Hull 1965, Atran 1990, voir la recension de Gelman et Hirschfeld 1999, Keil et Richardson 1999).<sup>6</sup>

---

<sup>4</sup> Ils écrivent (1999, 271) : « the place holder notion may overlook a related set of beliefs for even as people commonly believe in essences without knowing any details about those essences, they still might prefer some sorts of future details to others. Thus, there might be a physicalist bias, whereby people prefer essences to be seen as objects or 'stuff' rather than kinds of processes. »

<sup>5</sup> Machery, E., et Faucher, L., « Races : Cultural constructs or products of our evolved cognition ? », manuscrit.

<sup>6</sup> L'universalité de l'EP reste un sujet de discussion : Jeyifous 1986, cité dans Keil 1989, Walker 1992, et les travaux de Gil-White 2001 suggèrent que l'EP est présent dans plusieurs cultures. Fodor 1998 et Malt (« Reflections on water (and « water ») », manuscrit) expriment des doutes sur l'universalité de l'EP.

Je présenterai plusoin certaines données empiriques en psychologie développementale. Pour l'instant, cependant, j'admets l'existence de l'essentialisme psychologique.

### *1.2. L'essence de l'eau, c'est d'être composée de molécules d'H<sub>2</sub>O*

Croyons-nous que l'eau possède une essence ? Et si oui, avons-nous des croyances précises à propos de la nature de cette essence ? Intuitivement, la réponse est affirmative : nous croyons que l'essence de l'eau, c'est d'être composée de molécules d'H<sub>2</sub>O. Cette affirmation s'accorde avec les fameuses expériences de pensée de Putnam (1975).

Pour défendre la thèse selon laquelle la référence des noms communs n'est pas déterminée par les croyances des individus, Putnam demande au lecteur d'imaginer un monde contrefactuel, appelé Terre Jumelle, dans lequel se trouve un liquide qui a les mêmes propriétés perceptuelles et fonctionnelles que le liquide appelé « eau » dans le monde actuel. Seule la structure moléculaire du liquide sur Terre Jumelle le différencie du liquide que nous appelons « eau » : le liquide appelé « eau » dans le monde actuel est fait de molécules dont la structure atomique est H<sub>2</sub>O, alors que le liquide sur Terre Jumelle a une autre structure moléculaire, appelée « XYZ ». La question est la suivante : étant donné la différence chimique entre les deux liquides, le liquide sur Terre Jumelle est-il de l'eau ? L'intuition de Putnam est claire : ce n'est pas de l'eau. La grande majorité des philosophes partagent cette intuition (mais voir Mellor 1978, Dupré 1981 et Chomsky 1995) – y compris les adversaires de la théorie sémantique de Putnam (Evans 1982, Jackson 1998 et plus généralement les avocats des sémantiques bidimensionnelles pour les énoncés contrefactuels). Par conséquent, les philosophes ont en grande majorité l'intuition que la structure moléculaire est une propriété nécessaire de l'eau : ce qui n'est pas composé de molécules dont la structure chimique est H<sub>2</sub>O ne peut être de l'eau.<sup>7</sup> Cela est compatible à la fois avec des croyances essentialistes à propos de l'eau et avec la croyance plus spécifique que l'essence de l'eau, c'est d'être faite de molécules dont la structure est H<sub>2</sub>O. Le quasi consensus entre philosophes ne montre bien sûr pas que les sujets croient que l'essence de l'eau, c'est sa constitution chimique. Les philosophes ne forment sûrement pas un échantillon représentatif.

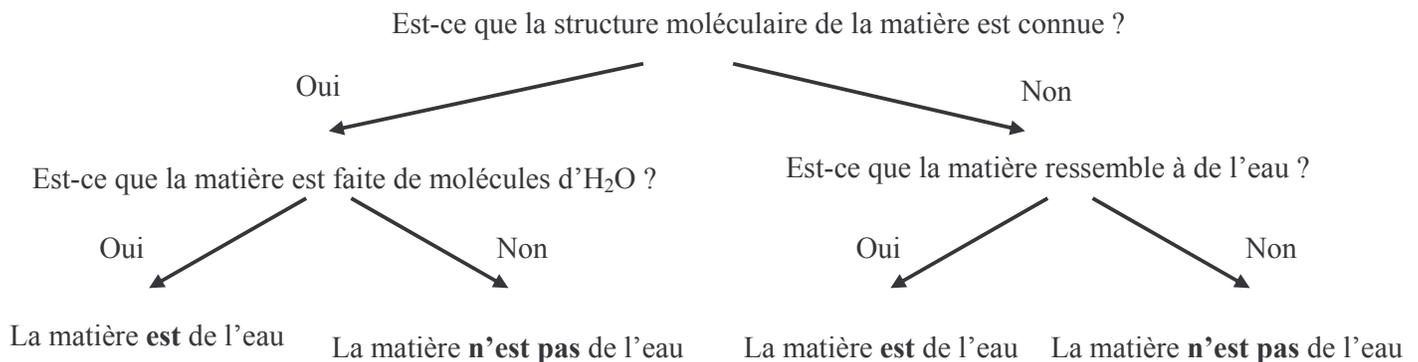
Supposons maintenant que nous soyons essentialistes à propos des matières et que nous croyions que l'essence de l'eau, c'est d'être composée de molécules dont la structure est

---

Norenzayan et al. 2003 avancent des données empiriques qui confirment ces doutes. Je ne développe pas ce point ici.

<sup>7</sup> De nombreux psychologues partagent les intuitions suscitées par ce type d'expériences de pensée (Keil 1989, 25-28, Keil 1991, 202-203, Gopnik et Meltzoff 1997, 162, Pinker et Prince 1999, 254-256, Prasada 1999, 132-134).

H<sub>2</sub>O. Comment catégoriserions-nous une matière comme étant de l'eau ? Il y a bien sûr deux possibilités. Nous connaissons la structure chimique de la matière que nous souhaitons catégoriser ou nous ne la connaissons pas. Si nous ne connaissons pas la structure moléculaire et ne pouvons pas la connaître, nous devons faire appel aux propriétés perceptuelles ou fonctionnelles de cette matière : si elle ressemble à de l'eau, si, par exemple, c'est un liquide transparent, inodore et sans goût, si elle est utilisée pour boire, se laver ou est située en des endroits particuliers, alors nous devrions conclure que c'est sans doute de l'eau. Au contraire, si nous connaissons la structure moléculaire de cette matière, nous devrions la catégoriser comme de l'eau si et seulement si elle est constituée de molécules dont la structure chimique est H<sub>2</sub>O. L'arbre de décision ci-dessous résume la procédure de catégorisation d'un essentialiste (voir, par exemple, Gil-White 2001, 519-520) :



**Schéma 7.1.**

Cette procédure prédit quand nous devons juger qu'un liquide est de l'eau. Est-ce que la procédure décrite ci-dessus correspond à l'usage de « eau » ? Pas toujours.

### 1.3. Comment utilisons-nous « eau » ?

La psychologue Barbara Malt a étudié en détail notre utilisation de « eau » (Malt 1993, 1994, manuscrit<sup>8</sup>). Ses expériences semblent montrer que les sujets ne suivent pas, en tout cas pas toujours, l'arbre de décision décrit ci-dessus.

La procédure expérimentale de la première expérience est la suivante (Malt 1994). Malt a collecté de nombreux exemples de liquides qui ne sont pas appelés « eau », mais qui, par certains aspects, ressemblent à de l'eau, par exemple, les larmes, le thé et le café. Elle a

<sup>8</sup> Malt, B. C., « Reflections on water (and « water ») », manuscrit.

aussi collecté de nombreux exemples de liquides qui sont habituellement appelés « eau », par exemple l'eau des lacs (« lake water »). La sélection de cette seconde classe de liquides repose sur l'observation de l'usage quotidien du mot « eau » (journaux, télévision...), sur une recherche par ordinateur dans un large corpus de mots (afin de trouver tous les liquides appelés « eau ») et sur une tâche expérimentale. Dans cette tâche, on demande aux sujets de donner tous les exemples possibles d'eau, des plus évidents aux plus obscurs. La sélection de la première classe de liquides repose sur l'étude des liquides vendus dans les supermarchés et sur une tâche expérimentale. Dans cette tâche, on demande aux sujets de donner des exemples de liquides qui ressemblent à de l'eau, mais qui ne sont pas appelés « eau ».

Dans la première expérience, on présente aux sujets les liquides appelés « eau » (voir ci-dessus) et on leur demande d'évaluer le pourcentage de molécules dont la structure chimique est  $H_2O$  dans ces liquides. On demande à un second groupe de sujets d'évaluer le pourcentage de molécules d' $H_2O$  dans les liquides qui ne sont pas appelés « eau », mais qui ressemblent à de l'eau.

Les résultats sont les suivants. Tout d'abord, les liquides qui sont habituellement appelés « eau » comme ceux qui ne sont pas habituellement appelés « eau » sont jugés être majoritairement composés de molécules d' $H_2O$  : selon les sujets, le pourcentage moyen d' $H_2O$  est de 67% dans les liquides qui ne sont pas appelés « eau » et de 83% dans les liquides qui sont appelés « eau ». Par exemple, le premier groupe juge en moyenne que l'eau pure est composée de 98,1% de molécules d' $H_2O$ , l'eau des mares de 78,8% de molécules d' $H_2O$  et l'eau des marais de 68,8%... Le second groupe juge en moyenne que le thé est composé de 91% de molécules d' $H_2O$ , la salive de 89,3% de molécules d' $H_2O$ , les larmes de 88,6% et la sueur de 87,3%...

Ensuite, de nombreux liquides qui ne sont pas appelés « eau » ont une proportion supérieure d' $H_2O$  que des liquides qui sont appelés « eau » : par exemple, le thé, le café, les larmes sont jugés avoir une proportion supérieure de molécules d' $H_2O$  que l'eau des mares ou l'eau des marais. En conséquence, la proportion d' $H_2O$  ne prédit pas quels liquides sont appelés « eau ». Avoir une proportion importante de molécules d' $H_2O$  n'est ni une condition suffisante (puisque les larmes ne sont pas dites être de l'eau) ni une condition nécessaire (puisque l'eau des radiateurs est dite être de l'eau) pour être appelé « eau ».

En résumé, *l'essentialisme psychologique semble faire une prédiction erronée à propos de l'utilisation de « eau »*. De prime abord, si nous étions essentialistes, nous devrions

juger qu'un liquide est de l'eau et être prêt à l'appeler « eau » si nous croyons qu'il est fait surtout de molécules d'H<sub>2</sub>O et nous devrions juger qu'un liquide n'est pas de l'eau et refuser de l'appeler « eau » si nous croyons qu'il contient peu de molécules d'H<sub>2</sub>O (voir l'arbre de décision du schéma 7.1.). Cependant, la proportion d'H<sub>2</sub>O n'explique pas quels liquides sont appelés « eau » : certains liquides comme le thé, la salive, le café, les larmes, la sueur ou la limonade ne sont pas dits être de l'eau, alors que les sujets croient qu'ils sont composés de davantage de molécules d'H<sub>2</sub>O que des liquides appelés « eau », comme l'eau des lacs ou l'eau des océans.<sup>9</sup>

Que faut-il en conclure ? De prime abord, la conclusion suivante s'impose : ou la thèse de l'essentialisme psychologique est fautive ou nous ne croyons pas que l'eau est par essence H<sub>2</sub>O. Nous verrons plus loin qu'il y a une autre explication de ces données.

#### *1.4. Les critiques d'Abbott*

Les résultats de Malt ont été discutés à de nombreuses reprises (Abbott 1997, 1999, Laporte 1998, Gelman et Hirschfeld 1999, Strevens 2000, Malt manuscrit, Bloom manuscrit<sup>10</sup>). Dans cette section, je présente les objections développées par Abbott.

Abbott soutient que les résultats de Malt disent en fait peu de choses sur nos jugements de catégorisation et sur notre disposition à appeler un liquide « eau ». Elle souligne à juste titre qu'appeler quelque chose un « x » dans une circonstance donnée et juger que cette chose est un x sont deux phénomènes tout à fait différents. Nous n'appelons pas toujours « x » certaines choses que nous jugeons être des x et que nous sommes disposés à appeler « x ». Les exemples suivants illustrent l'esprit de cette remarque :

- (1) Si l'on pointait vers un chien et si l'on demandait à quelqu'un ce que c'est, il répondrait probablement que c'est un chien ou un rottweiler, mais il ne répondrait pas que c'est un animal. Cependant, il jugerait de toute évidence que c'est un animal et serait prêt à l'appeler « animal ».
- (2) Si l'on pointait vers un singe rhésus et si l'on demandait à un primatologue ce que c'est, il dirait sans doute que c'est un singe rhésus, mais sûrement pas que c'est un Macaca Mulatta. Cependant, il jugerait probablement que c'est bien un Macaca Mulatta et serait prêt à l'appeler « Macaca Mulatta ».

---

<sup>9</sup> Voir aussi Chomsky 1995.

<sup>10</sup> Malt, B. C., « Reflections on water (and « water ») », manuscrit, Bloom, P., « Water as an artifact kind », manuscrit.

Ces exemples montrent bien qu'appeler quelque chose un x et juger que cette chose est un x ne sont pas des phénomènes identiques. Dans de nombreuses circonstances, nous n'appelons pas quelque chose un x, alors même que nous jugeons que c'est un x et que nous sommes disposés à le nommer « x ».

La manière dont un locuteur appelle un objet dépend pour l'essentiel de la situation communicative entre ce locuteur et l'auditoire. Ainsi, bien qu'un locuteur puisse être convaincu qu'un objet est un x, il peut habituellement éviter de l'appeler « x » pour des raisons pragmatiques. Une telle situation peut avoir différentes causes. Le nom peut être trop technique (voir l'exemple 2) et peut donc être inapproprié pour de nombreux auditeurs. Le locuteur peut aussi connaître le nom d'une classe moins large à laquelle l'objet appartient (voir l'exemple 1) : puisque référer à cette classe est plus informatif que référer à une classe plus large, sans être moins pertinent, le locuteur a tendance à éviter le nom de la classe plus large (voir la première maxime de quantité dans Grice 1989, 26, voir aussi Sperber et Wilson 1986).

Abbott suggère que cette distinction explique les résultats de Malt. Elle concède l'existence d'une dissociation entre les liquides qui sont habituellement *nommés* « eau » et les croyances sur la constitution chimique de ces liquides. Cependant, selon elle, cela ne montre pas qu'il existe une dissociation entre, d'une part, les liquides qui sont jugés *être* de l'eau et que nous sommes prêts à appeler « eau » et, d'autre part, les croyances sur la constitution chimique de ces liquides. Nous croyons qu'un liquide est de l'eau et nous sommes disposés à l'appeler « eau » si et seulement s'il est composé pour l'essentiel de molécules dont la structure chimique est H<sub>2</sub>O, même si, pour des raisons pragmatiques, nous n'appelons pas toujours « eau » ces liquides.<sup>11</sup>

Elle offre deux arguments en faveur de cette thèse. D'abord, le thé, les larmes, le café etc. ont des noms spécifiques, respectivement « thé », « larmes » et « café ». Puisque « eau » est moins spécifique que ces noms, les locuteurs doivent appeler ces liquides « thé », « larmes » et « café » plutôt que « eau » (première maxime de la quantité), même s'ils croient que ce sont des eaux et sont disposés à les appeler ainsi. Abbott soutient aussi que ces noms sont associés à des éléments descriptifs : par exemple, « larmes » peut être associé à la description « gouttes qui sortent des yeux lorsque l'on est triste ». Par conséquent, ces mots

---

<sup>11</sup> Abbott écrit ainsi à propos des larmes et de l'eau des océans (Abbott 1997, 315) : « The difference is not that one substance (e.g., the stuff coming out of your eyes) is not water and the other stuff is water. They are both water. The difference is only that one is *called* 'water' and the other is not. And that only goes to show that we need to distinguish what something is from what it may be called on particular occasions. »

peuvent correspondre davantage aux intentions communicatives des locuteurs que le terme « eau ». Elle conclut que l'utilisation de « eau » ne montre pas que nous doutions que le thé, les larmes et le café soient des eaux.

### *1.5. Réponse aux critiques de Abbott*

Je ne trouve pas les critiques de Abbott convaincantes (voir aussi Laporte 1998, Strevens 2000, Malt manuscrit, Bloom manuscrit<sup>12</sup>). De toute évidence, la manière dont nous utilisons « eau » est affectée par les principes pragmatiques de la communication. Par exemple, l'eau est toujours plus ou moins pure. Le degré d'impureté en dessous duquel un liquide n'est pas *appelé* « eau » est sans doute déterminé pragmatiquement, par exemple en fonction des classes de contraste saillantes. Cependant, je soutiens que nous ne croyons pas que le thé, le café etc. soient de l'eau.<sup>13</sup> De nombreuses objections ont été avancées contre Abbott (voir les références mentionnées ci-dessus). Je propose l'objection suivante.

Les cas (1) et (2) mentionnés ci-dessus sont très différents des cas du thé et du café. Dans les cas (1) et (2), deux expressions linguistiques peuvent être appliquées à un même objet. Pour des raisons pragmatiques, les locuteurs utilisent habituellement l'une d'entre elles. Cependant, ils sont prêts à approuver l'utilisation de l'autre expression – au moins dans certaines circonstances linguistiques. Considérons par exemple les questions suivantes (posées à propos des objets concernés par (1) et (2)) :

(3) Est-ce que c'est un animal ?

(4) Est-ce que c'est un Macaca Mulatta ?

Les réponses seraient sans doute affirmatives. En raison de la forme interrogative de (3) et (4), la règle pragmatique selon laquelle toutes choses étant égales, un locuteur doit utiliser l'expression la plus informative n'exclut pas l'usage de « animal » et de « Macaca Mulatta ». Cependant, comparons (3) et (4) à la question suivante :

(5) Est-ce que cela est de l'eau ?

---

<sup>12</sup> Malt, B. C., « Reflections on water (and « water ») », manuscrit, Bloom, P., « Water as an artifact kind », manuscrit.

<sup>13</sup> Il va de soi que nous croyons que le thé est composé d'eau ou fait avec de l'eau. Ce n'est pas la question.

où « cela » réfère à du thé, à du café ou à des larmes. Je pense que la réponse à cette question serait négative.

Admettons que les individus répondent différemment à (3)-(4) et à (5). Si les résultats présentés par Malt pouvaient être expliqués comme le suggère Abbott, une telle différence serait incompréhensible. La maxime de quantité explique pourquoi nous utilisons « chien » plutôt que « animal » pour parler de Médor. Si elle expliquait aussi pourquoi nous n'utilisons pas « eau » pour parler du thé ou du sang, la différence entre (3)-(4) et (5) serait mystérieuse. La deuxième raison avancée par Abbott, à savoir l'association de « thé », « sang »... avec des descriptions, n'explique pas non plus notre réponse négative à (5). Les deux arguments avancés par Malt ne sont donc pas convaincants. Cette réponse négative montre au contraire que nous croyons que le thé, le café, les larmes etc. ne sont pas de l'eau et que nous ne sommes pas disposés à les appeler « eau ». Je conclus donc que l'interprétation des résultats de Malt proposée par Abbott est sans doute fausse.

### *1.6. Conclusion*

Les données empiriques sur le phénomène psychologique appelé « essentialisme psychologique »<sup>14</sup> et les résultats de Malt recensés dans cette section établissent l'existence d'une dissociation entre les croyances essentialistes sur certaines catégories et l'utilisation des noms correspondants :

- nous croyons que certaines catégories (au moins les catégories d'animaux et de plantes) et les matières sont définies par une essence.
- nous utilisons souvent les noms de ces catégories et de ces matières indépendamment de cette croyance.

L'eau illustre cette dissociation : dans les exemples proposés par Malt, la croyance que l'eau est essentiellement composée de molécules dont la structure chimique est H<sub>2</sub>O ne détermine pas l'usage du nom « eau ». Malt écrit ainsi (1993, 384) :

*« [people] may hold beliefs about how categories are constrained and about how they themselves make classification judgments, that do not accurately reflect the actual constraints or judgment processes. Many factors may give rise to erroneous naive*

---

<sup>14</sup> Je vais présenter certaines données dans la section suivante.

beliefs about categorization, including incomplete introspective access to the relevant representations and processes, and the tendency to retrieve mainly prototypes and clear-cut cases of category members when thinking about category membership. »<sup>15</sup>

Abbott a suggéré que les résultats présentés par Malt pouvaient être expliqués en termes pragmatiques : pour des raisons pragmatiques, essentiellement en vertu de la première maxime de quantité de Grice, nous n'appelons pas le thé, le café, la limonade, le sang etc., de l'eau, alors même que nous croyons que ce sont de l'eau. En conséquence, la dissociation entre nos intuitions essentialistes et nos pratiques de nomination n'est que superficielle. Je rejette cependant cette interprétation des résultats présentés par Malt, car notre réticence à appeler le thé, le café etc. de l'eau ne semble pas résulter d'une règle pragmatique.

Comment comprendre cette dissociation ? Le reste de ce chapitre s'efforce de répondre à cette question.

## 2. Avons-nous un concept essentialiste d'eau ?

Dans cette section, je présente et critique la manière dont Malt interprète ses propres données.

### 2.1. L'interprétation proposée par Malt

Comment Malt interprète-t-elle la dissociation entre notre croyance à propos de l'essence de l'eau (et notre approbation de l'énoncé « l'eau, c'est H<sub>2</sub>O ») et certaines pratiques de dénomination ? Selon elle, cette dissociation montre que *notre concept* d'eau ne spécifie pas que l'eau est essentiellement constituée de molécules dont la constitution chimique est H<sub>2</sub>O. Elle admet que nous *croyons* que l'eau est essentiellement constituée de molécules dont la constitution chimique est H<sub>2</sub>O, mais cette croyance ne fait pas partie de notre concept d'eau, c'est-à-dire de l'ensemble de connaissances que nous utilisons par défaut dans nos processus cognitifs à propos de l'eau. Ainsi, selon Malt, nous avons un concept non essentialiste d'eau, qui explique notre pratique de dénomination, et une croyance essentialiste que nous rapportons quand on nous demande ce qu'est vraiment l'eau. Elle suggère que cette croyance essentialiste résulte en partie de notre éducation dans une société où la science occupe une place centrale (voir aussi Fodor 1998).

Quelle est la différence entre soutenir que nous avons un concept essentialiste d'eau et soutenir que nous croyons simplement que l'eau est par essence constituée de molécules

---

<sup>15</sup> Je souligne.

d'H<sub>2</sub>O ? La différence est la suivante (chapitre 1, schéma 1.4.) : les concepts sont des bases de données (« data structure ») qui sont utilisées *par défaut* dans nos processus cognitifs alors que nos simples croyances sont utilisées *occasionnellement*. Par exemple, l'ensemble de connaissances à propos des grands-mères qui forme notre concept de grand-mère est utilisé par défaut quand nous raisonnons à propos des grand-mères, tandis que les croyances qui ne font pas partie de notre concept sont utilisées occasionnellement. La question est donc : notre croyance essentialiste à propos de l'eau est-elle utilisée par défaut dans nos processus cognitifs ou simplement occasionnellement ?

## 2.2. Données en faveur de l'interprétation de Malt

Le rejet par Malt des concepts essentialistes présuppose que la croyance essentialiste que l'eau est faite de molécules d'H<sub>2</sub>O – ou d'autres croyances de ce type (par exemple, que les chiens ont une structure génétique particulière, ont des chiens pour parents et engendrent eux-mêmes des chiens...) – n'est pas activée par défaut dans nos processus cognitifs à propos de l'eau, par exemple quand nous faisons une analogie entre l'eau et quelque chose d'autre.

Les résultats de Malt semblent venir à l'appui de cette thèse : notre pratique de dénomination (application de « eau » à des liquides) ne semble pas reposer sur notre croyance essentialiste. Cette thèse est aussi compatible avec de nombreuses données empiriques.

Smith et Sloman ont reproduit l'expérience de la pizza conduite par Rips (Rips 1989, Smith et Sloman 1994, voir le chapitre 5, section 3.1.2.). Les sujets lisent la description d'un objet rond et doivent décider si c'est une pizza ou une pièce américaine de 25 cents. Dans une condition, les sujets connaissent seulement le diamètre de l'objet (comme dans l'expérience de Rips). Dans l'autre, les sujets connaissent le diamètre de l'objet *et* certaines autres propriétés, par exemple le fait qu'il est argenté. Cette couleur est bien sûr typique des pièces de 25 cents, quoiqu'elle ne soit pas nécessaire : on peut recouvrir les pièces de 25 cents de peinture. Puisque le diamètre des pièces de 25 cents est fixé par la loi, c'est une propriété nécessaire de ces pièces. En conséquence, les sujets pourraient appliquer dans les deux conditions la règle : « si un objet n'a pas la taille *c*, ce n'est pas une pièce de 25 cents » (voir la section 1., schéma 7.1. sur la procédure de décision liée à l'essentialisme psychologique). Quand on leur demande de ne pas raisonner à voix haute, les sujets n'appliquent pas cette règle (expérience 1). Quand on leur demande de raisonner à voix haute, ils appliquent cette règle seulement dans la première condition, c'est-à-dire quand ils connaissent seulement le diamètre de l'objet cible. S'ils savent aussi que l'objet a la couleur typique des pièces de 25

cents, ils négligent le fait que le diamètre est une propriété nécessaire de ces pièces (expérience 2).

Ces résultats pourraient être interprétés comme suit (mais voir ci-dessous) : nous ne pensons pas par défaut aux pièces de 25 cents comme ayant une propriété nécessaire, à savoir un diamètre fixé par la loi. Sinon, les sujets appliqueraient la règle « si un objet n'a pas la taille *c*, ce n'est pas une pièce de 25 cents » et ils ne prendraient pas en compte la couleur de l'objet cible. Au contraire, les sujets semblent décider si quelque chose est une pièce de 25 cents sur la base des propriétés typiques de ces pièces. En conséquence, notre concept de pièces de 25 cents semble être un prototype : quand nous devons prendre une décision à propos des pièces de 25 cents, nous activons en MT les propriétés typiques de ces pièces.<sup>16</sup> Nous pouvons *croire* qu'elles doivent avoir un diamètre spécifique, mais cette connaissance n'est pas activée par défaut quand nous raisonnons à propos des pièces de 25 cents, par exemple quand nous décidons si quelque chose est une pièce de 25 cents. Cette croyance ne semble donc pas faire partie de notre *concept* de pièces de 25 cents.

Les résultats Hampton (1995) s'accordent aussi avec ceux de Malt. L'expérience est semblable à l'expérience 2 de Smith et Sloman 1994. Il s'agit de déterminer si les décisions de catégorisation sont affectées par la présence ou l'absence des propriétés caractéristiques d'une catégorie. On présente aux sujets la description d'un objet. La tâche est de décider si cet objet appartient à une catégorie donnée. On demande explicitement aux sujets de prendre en compte la définition réelle de la catégorie concernée. Le principal résultat est le suivant : pour presque toutes les catégories, il fut impossible de trouver un ensemble de propriétés traitées par les sujets comme étant nécessaires et suffisantes pour appartenir à ces catégories. La présence de propriétés typiques influence toujours les décisions de catégorisation. Cela est aussi bien le cas pour les catégories d'artéfacts que pour les catégories d'animaux, plantes etc. Un exemple peut éclairer cette expérience. Hampton utilise la description suivante :

« The offspring of two zebras. The creature was given a special nutritional diet during development. It now looks and behaves just like a horse, with a uniform brown color. »

---

<sup>16</sup> Ou bien un ensemble d'exemplaires.

Deux tiers des sujets soutiennent que c'est un cheval, bien que ses parents soient des zèbres. Ce résultat est difficilement compatible avec l'existence d'un concept essentialiste de zèbre qui spécifierait que les zèbres engendrent des zèbres (Rips 1989, Gelman et Wellman 1991). La quasi impossibilité de trouver des propriétés qui soient traitées comme étant nécessaires et suffisantes dans des tâches de catégorisation supportent la thèse de Malt : nous avons des croyances essentialistes, mais pas de concepts essentialistes.<sup>17</sup>

Ces données sont sans aucun doute suggestives. Elles montrent que nous avons des concepts non essentialistes d'espèces naturelles et d'artéfacts. Elles montrent aussi que ces concepts sont facilement accessibles et sont utilisés dans des tâches de catégorisation, alors même que les sujets pourraient prendre leur décision de catégorisation sur la base de propriétés nécessaires. Mais montrent-elles vraiment que nous n'avons pas de concepts essentialistes ? De toute évidence, la thèse selon laquelle nous n'avons pas de concepts essentialistes prédit les résultats de Malt, Smith et Sloman et Hampton. Cependant, ces résultats sont aussi explicables par une autre hypothèse. Nous avons bien des concepts essentialistes. Mais ces concepts essentialistes ne sont pas activés par défaut dans *certaines* tâches de catégorisation. Ils sont activés régulièrement dans *d'autres* contextes. Ni l'expérience de Smith et Sloman ni l'expérience de Hampton ne les activent. On pourrait même soutenir que les expériences de Smith et Sloman montrent en fait que dans certains contextes (verbalisation, rareté des indices), nous utilisons par défaut des concepts essentialistes.<sup>18</sup> J'examine maintenant cette hypothèse.<sup>19</sup>

### 2.3. Les données contre l'interprétation de Malt

En fait, de très nombreuses données suggèrent que nous avons des *concepts* essentialistes, c'est-à-dire des croyances essentialistes qui sont utilisées par défaut dans nos processus cognitifs (Keil 1989, voir aussi Keil et Batterman 1984, Keil 1987, Gelman et Markman 1986, 1987, Gelman 1988, Gelman et Wellman 1991, Gelman et al. 1994). Je présente maintenant

---

<sup>17</sup> Malt et Johnson (1992) ont aussi montré que ni la forme ni la fonction des artéfacts ne sont traitées comme l'essence des artéfacts dans des tâches de catégorisation.

<sup>18</sup> Par ailleurs, les résultats de Hampton suggèrent que l'accessibilité des concepts essentialistes peut varier d'un individu à l'autre.

<sup>19</sup> On pourrait aussi proposer que les deux concepts sont activés en parallèle, même si un seul détermine les jugements des sujets.

les travaux importants de Frank Keil sur le développement des performances de catégorisation dans des espèces naturelles (Keil 1989).<sup>20</sup>

Ces données portent surtout sur la question suivante : est-ce que les adultes et les enfants modifient leur catégorisation d'un objet, quand certaines propriétés de cet objet sont modifiées ? Le principal résultat de cette recherche est que même de très jeunes enfants (dès trois ans avec certaines procédures expérimentales – voir Keil 1989, chapitre 11) ne modifient pas la catégorie d'un objet naturel (animal, plante etc.), quand son apparence physique est modifiée. Les adultes font de même. Ces données supportent fortement l'idée d'un développement précoce de l'essentialisme psychologique (mais voir Strevens 2000 et 2001).

Considérons plus en détail les données expérimentales. L'une des procédures expérimentales de Keil, appelée habituellement la procédure de transformation (« transformation design »), est la suivante.<sup>21</sup> Les enfants lisent une histoire à propos d'un objet. Sa catégorie est explicitement spécifiée. L'histoire explique que l'apparence physique de cet objet est modifiée par un processus particulier. Au terme de ce processus, l'objet ressemble aux membres d'une autre catégorie (la catégorie finale). La nature de ce processus est clairement explicitée. Les enfants doivent décider si l'objet appartient toujours à sa catégorie originelle ou s'il appartient désormais à la catégorie finale. Par exemple, dans l'expérience célèbre du cheval et du zèbre, les enfants lisent l'histoire suivante (Keil 1989, 307) :

« The doctors took a horse and did an operation that put black and white stripes all over its body. They cut off its mane and braided its tail. They trained it to stop neighing like a horse, and they trained it to eat wild grass instead of oats and hay. They also trained it to live in the wilds in Africa instead of in a stable. When they were all done, the animal looked just like this. When they were finished, was this animal a horse or a zebra? »

Les expériences de Keil mesurent l'effet de plusieurs variables indépendantes. L'objet appartient à différents types de catégories (espèces naturelles ou catégories d'artéfacts). En outre, les catégories originelles et finales peuvent être plus ou moins différentes. Finalement,

---

<sup>20</sup> Au sens psychologique du terme – voir le chapitre 2, section 1.1.. Dans ce qui suit, « espèce naturelle » est utilisé en ce sens.

<sup>21</sup> L'autre procédure est la procédure de découverte (« the discovery design ») : des chercheurs découvrent qu'un animal qui possède les propriétés physiques typiques d'une espèce, par exemple des chiens, a en fait des organes internes d'une autre espèce, par exemple des chats.

différents processus, plus ou moins superficiels (voir ci-dessous), peuvent affecter l'apparence des objets (Keil 1989, chapitre 11).

Les résultats sont les suivants.

- Les espèces naturelles et les artefacts sont traités différemment. Les adultes et, pour la plupart des modifications, les enfants de 10 ans croient que la modification de l'apparence physique et du comportement d'un membre d'une espèce naturelle ne modifie pas son appartenance catégorielle. Au contraire, dès six ans, les enfants croient que la modification de l'apparence physique et de la fonction d'un artefact modifie son appartenance catégorielle : il appartient à la catégorie finale.
- Le type de modification affecte l'âge auquel les enfants manifestent le comportement décrit ci-dessus pour les espèces naturelles. Plus le processus de modification est superficiel, plus tôt les enfants manifestent ce comportement.
  - La modification peut être due à un costume. Par exemple, dans une histoire, un costume de zèbre est mis sur un cheval. Même des enfants de trois ans disent que le cheval n'est pas devenu un zèbre.
  - Elle peut être due à une modification cosmétique temporaire (maquillage, peinture...). Les enfants de six ans négligent, en règle générale, l'effet de ces modifications temporaires. Cependant, un certain nombre ne les négligent pas. Ils soutiennent par exemple qu'un raton laveur qui est rasé, peint en noir et blanc et auquel on implante une poche émettant une odeur nauséabonde est un putois. Les enfants de huit ans négligent toujours ce type de modifications.
  - La modification cosmétique peut être permanente. Les enfants de dix ans négligent d'habitude cette modification.
  - La modification peut avoir un caractère médical (pilules, injection...). Les enfants de dix ans ne négligent pas systématiquement les modifications induites par ce type de processus. Les adultes négligent toutes les modifications.
- La distance entre la catégorie originelle et la catégorie finale affecte les réponses des enfants. Quand les deux catégories appartiennent à des domaines différents (par exemple, une plante et un animal), même des enfants de cinq ans répondent que l'identité de l'animal n'est pas modifiée par un processus de type biologique.

*Que montrent ces résultats ?* Tout d'abord, les adultes et les enfants de dix ans semblent croire que l'apparence physique ne détermine pas l'appartenance à une espèce naturelle, par exemple à une catégorie d'animaux ou à une catégorie de plantes.<sup>22</sup> Cela est compatible avec l'idée que les enfants acquièrent tôt une croyance essentialiste à propos de l'identité des espèces naturelles : si les enfants de dix ans et les adultes croient que l'identité d'une espèce naturelle est constituée par une essence, alors modifier l'apparence des objets n'implique aucun changement catégoriel. En outre, ces expériences ainsi que les autres travaux sur l'essentialisme psychologique montrent que nos croyances essentialistes forment des concepts. Elles sont utilisées par défaut dans *certaines* tâches cognitives.

Le rejet par Malt des concepts essentialistes s'accorde mal avec les résultats de Keil. Comment Malt explique-t-elle ces résultats sans attribuer aux enfants des concepts essentialistes ? Elle a suggéré la réponse suivante (Malt 1993). Selon elle, les performances des enfants résultent sans doute des particularités des expériences de Keil.<sup>23</sup> Cependant, cette idée est très peu probable. Car Keil a utilisé deux procédures expérimentales, la procédure de transformation et la procédure de découverte. Les résultats de ces deux procédures expérimentales convergent : les enfants semblent suivre la règle selon laquelle, si des animaux ont les organes des x à l'intérieur de leur corps, s'ils ont été engendrés par des x et s'ils engendrent des x, alors ce *sont* des x, même s'ils ressemblent à des y. Les concepts de ces animaux semblent donc exprimer la règle suivante : si une entité a été engendrée par des x, si elle engendre des x et si elle a les organes des x à l'intérieur du corps, c'est un x. Cette convergence suggère que l'objection de Malt n'est pas crédible : quand deux procédures expérimentales convergent sur les mêmes données, ces données ne résultent pas des particularités d'une procédure particulière.

Je conclus que son rejet des concepts essentialistes n'est pas crédible. Nous croyons, au moins dans certains domaines (voir la section 1.1. ci-dessus), que les catégories ont des essences. Malt accepte cette thèse. Les résultats de Keil montrent que dans certains contextes, cette croyance est utilisée par défaut quand nous raisonnons à propos de l'appartenance catégorielle. Ces croyances forment donc des concepts essentialistes (voir la section 2.1. ci-dessus).

#### *2.4. Quand ces différents concepts sont-ils activés en MT ?*

---

<sup>22</sup> Cela ne montre pas qu'ils tiennent l'apparence d'un objet pour dépourvue de pertinence. L'apparence physique d'un objet reste un indice sur sa catégorie.

<sup>23</sup> Je ne sais pas si elle soutient toujours ce point de vue. Elle n'aborde pas cette question dans Malt, B. C., « Reflections on water (and « water ») », manuscrit.

Supposons que nous ayons plusieurs concepts d'eau (et de chats, et, sans doute, de tables<sup>24</sup> etc.). Cela soulève alors le problème suivant : quand ces différents concepts sont-ils activés ? Quand pensons-nous à l'eau de manière essentialiste (en extrayant EAU<sub>E</sub> de la MLT) et quand pensons-nous à l'eau de manière non essentialiste (en extrayant un concept non essentialiste d'eau de la MLT) ?

Bien sûr, cette question est empirique. Cependant, les données présentées ci-dessus invitent aux spéculations suivantes.

- Tout d'abord, notre concept non essentialiste d'eau (ou nos concepts non essentialistes d'eau – voir la section suivante) semble être activé par défaut quand nous nous engageons dans certaines tâches de catégorisation. Les résultats de Smith et Sloman (1994) montrent que dans de nombreuses situations, nos décisions de catégorisation ne reposent pas sur notre croyance en des essences. Les résultats de Hampton montrent pareillement qu'il est souvent difficile d'inviter les sujets à traiter certaines propriétés comme nécessaires, quand ils doivent décider si quelque chose appartient à une catégorie. Cela est vrai aussi bien des artéfacts que des espèces naturelles.
- Les résultats de Sloman montrent aussi que le raisonnement verbal à voix haute favorise l'activation des concepts essentialistes (voir aussi Malt 1993 et le chapitre 5, section 3.1.). Quand les sujets doivent raisonner à voix haute, ils traitent souvent certaines propriétés comme étant essentielles et ils peuvent faire reposer leurs décisions de catégorisation sur ces propriétés.
- Les données de Keil montrent que les concepts essentialistes peuvent être activés quand les décisions de catégorisation sont problématiques. Dans les expériences de découverte et de transformation, on présente aux enfants et aux adultes plusieurs informations qui suggèrent des jugements opposés. Ils doivent prendre en compte l'ensemble de ces informations afin de catégoriser les cibles. Il se peut que quand nous avons à prendre en compte des informations incompatibles, nous fassions plus facilement appel à des concepts essentialistes.

## 2.5. Conclusion

Dans cette section, j'ai critiqué l'interprétation que Malt propose de ses propres résultats. Malt soutient que nous n'avons pas de concepts essentialistes, mais plutôt des croyances

---

<sup>24</sup> L'essence d'une table pourrait être fonctionnelle ou intentionnelle (Rips 1989, Bloom 1996, mais voir Malt et Johnson 1992) : une table est un objet qui est utilisé pour supporter des objets etc. ou une table est un objet qui a été produit en vue de supporter des objets etc.

essentialistes que nous rapportons quand on nous questionne sur la nature des catégories. Ces croyances jouent un rôle limité dans notre cognition. De nombreuses données sont compatibles avec cette interprétation. Cependant, il faut rejeter cette interprétation. Des travaux expérimentaux importants, en premier lieu les travaux de Keil, montrent que nous avons bien des concepts essentialistes : dans certains contextes, nos croyances essentialistes sont activées par défaut.

### 3. EAU<sub>E</sub> et EAU<sub>NE</sub>

Nous avons vu qu'il existe au moins deux concepts de l'eau, un concept essentialiste et un concept non essentialiste. Quelle la nature de ces deux concepts ? Cette section s'efforce de répondre à cette question. Contre Bloom (ms) et Gelman et Hirschfeld (1999), je soutiens que le concept essentialiste est un concept théorique, tandis que le concept non essentialiste est un prototype ou un ensemble d'exemplaires.

#### 3.1. *L'interprétation proposée par Bloom*

Comme Abbott, le psychologue Paul Bloom admet les données de Malt (Bloom<sup>25</sup>). Il suggère de les interpréter comme suit. Il y a deux concepts d'eau. L'un est essentialiste (EAU<sub>E</sub>) et l'autre ne l'est pas (EAU<sub>NE</sub>). Autrement dit, il y a deux manières de penser à l'eau. Nous pouvons penser à l'eau de manière essentialiste : quelque chose est de l'eau si et seulement s'il est constitué de molécules dont la structure atomique est H<sub>2</sub>O (si l'on suppose que nous croyons que H<sub>2</sub>O est l'essence de l'eau). Mais nous pouvons aussi penser à l'eau de manière non essentialiste : un liquide est de l'eau s'il est surtout composé de molécules d'H<sub>2</sub>O, est utilisé d'une manière spécifique et si on le trouve en des endroits spécifiques. Aucune de ces propriétés n'est tenue pour nécessaire.

Les expériences de Malt montrent que nous avons un concept non essentialiste d'eau. Pour défendre l'existence d'un concept essentialiste d'eau, Bloom cite nos intuitions à propos des conditions de vérité des énoncés suivants :

(6) L'eau, c'est H<sub>2</sub>O (« water is H<sub>2</sub>O »)

(7) Les vers de terre, les méduses, les poulets et les bébés sont composés pour l'essentiel d'eau.

---

<sup>25</sup> Bloom, P., « Water as an artifact kind », manuscrit.

Nous croyons que l'énoncé (6) est vrai. Par ailleurs, les conditions de vérité de l'énoncé (7) ne mentionnent pas l'origine, l'utilisation et la fonction du liquide dans les vers de terre, les méduses, les poulets et les bébés, mais seulement sa composition chimique : (7) est vrai si et seulement si les vers de terre, les méduses, les poulets et les bébés sont composés pour l'essentiel de molécules d'H<sub>2</sub>O. Cela suggère que nous avons un concept essentialiste d'eau qui est à l'occasion exprimé par « eau ».<sup>26</sup>

Quelle est la nature de ces deux concepts ? Bloom développe une idée de Frank Keil (1994). Keil a soutenu que nous avons différents modes d'interprétation de la réalité (« modes of construal ») – en termes d'artéfacts, en termes d'espèces naturelles... Bloom propose que nous pouvons parfois penser à une même catégorie de différentes manières, par exemple comme étant une classe d'artéfacts et comme étant une espèce naturelle.<sup>27</sup> Par exemple, il est possible de penser au Sprite comme étant une matière naturelle ou comme étant un artéfact. Dans le premier cas, le Sprite est identifié à sa structure chimique : s'il arrivait qu'un liquide avec cette même structure chimique soit un remède efficace contre la calvitie masculine, et si on en produisait à cette fin, on pourrait alors dire que le Sprite sert à lutter contre la calvitie. Nous pourrions toujours utiliser le terme « Sprite » si la fonction du liquide avait entièrement changé. Notons que dans ce cas-là, le Sprite est identifié à sa composition chimique : un liquide est du Sprite si et seulement s'il a une structure chimique spécifique. Dans le second cas, le Sprite serait caractérisé par son origine intentionnelle. La compagnie qui produit le Sprite pourrait en modifier la structure chimique. Si le liquide résultant de cette modification avait un goût et une apparence semblables au Sprite actuel, ce serait toujours du Sprite (cela fut le cas du coca-cola) : nous pourrions toujours utiliser le terme « Sprite » si la structure chimique du liquide avait entièrement changé. Bloom soutient que le sang peut pareillement être caractérisé essentiellement ou fonctionnellement : un liquide est du sang si et seulement s'il a une certaine structure chimique ou si et seulement s'il a une certaine fonction dans les organismes. Bloom affirme que cela est aussi le cas de « eau » : ce nom est associé avec un concept qui résulte d'un mode d'interprétation de la réalité en termes d'artéfacts et avec un concept qui résulte d'un mode d'interprétation de la réalité en termes d'espèces naturelles. Le second concept est un concept essentialiste, EAU<sub>E</sub>, puisque le mode d'interprétation en termes d'espèces naturelles est essentialiste. Le second, EAU<sub>NE</sub>, n'est pas essentialiste puisque, selon Bloom, l'essentialisme ne fait pas partie de notre appréhension de la réalité en

---

<sup>26</sup> Cet argument s'ajoute à l'argument proposé ci-dessus à partir des travaux de Keil.

<sup>27</sup> « espèce naturelle » réfère aussi aux matières comme l'or, l'eau etc.

termes d'artéfacts. Les travaux de Malt portent uniquement sur l'utilisation du second concept.

Remarquons que selon Bloom, les *deux* concepts d'eau sont *théoriques*. EAU<sub>E</sub> est un concept théorique d'eau en tant qu'espèce naturelle, tandis que EAU<sub>NE</sub> est un concept théorique d'eau en tant qu'artéfact. Ces deux concepts encodent des propositions théoriques.

### 3.2. L'interprétation de Gelman et Hirschfeld

Gelman et Hirschfeld ont avancé une autre interprétation des résultats de Malt (Gelman et Hirschfeld 1999, 407-408).<sup>28</sup> Comme Bloom, ils soutiennent que nous avons deux concepts d'eau, un concept essentialiste et un concept non essentialiste. Cependant, ils conçoivent la nature de ces deux concepts différemment. Le concept non essentialiste d'eau est un concept générique : il réfère à un grand nombre de liquides, parmi lesquels se trouve le liquide auquel réfère le concept essentialiste d'eau. Le schéma ci-dessous résume cette proposition :

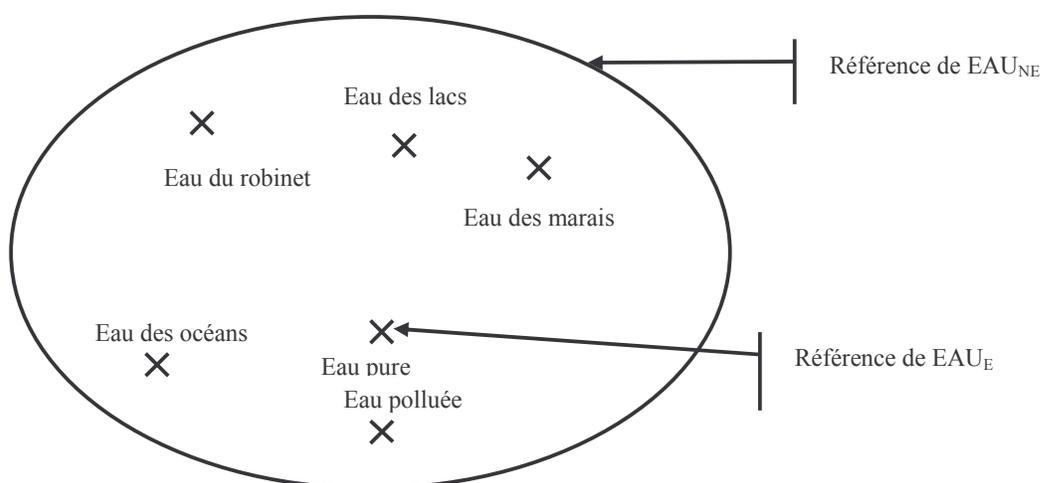


Schéma 7.2.

« eau » est donc ambigu. Il peut référer ou bien à de l'eau pure ou bien à une classe plus large de liquides. En conséquence, on peut juger que les deux énoncés suivants sont vrais :

(8) Le liquide dans la Seine est de l'eau

<sup>28</sup> Dans la discussion qui suivit un exposé de S. Gelman, P. Bloom a approuvé cette approche (Maryland Workshop on Innateness and the Structure of the Mind, organisé par P. Carruthers, mars 2002). Il n'est cependant pas certain que les idées de Gelman et Hirschfeld soient compatibles avec la proposition de Bloom présentée ci-dessus. Si EAU<sub>E</sub> et EAU<sub>NE</sub> résultent de différents modes d'interprétation de la réalité, pourquoi l'un de ces deux concepts serait-il générique et l'autre spécifique ?

(9) Le liquide dans la Seine n'est pas de l'eau.

(8) et (9) sont l'un et l'autre vrais parce que « eau » exprime deux concepts différents dans ces deux énoncés.

### 3.3. *Quelle est la nature de EAU<sub>E</sub> et EAU<sub>NE</sub> ?*

Comme Bloom, Hirschfeld et Gelman, je soutiens que nous avons plusieurs concepts d'eau : nous avons bien un concept essentialiste et (au moins) un concept non essentialiste d'eau. Cependant, je n'approuve pas la manière dont ils caractérisent EAU<sub>NE</sub>. Bloom propose en particulier que c'est un concept théorique. J'affirme que c'est un prototype ou un ensemble d'exemplaires. Par ailleurs, je soutiens que nous avons sans doute plusieurs concepts non essentialistes d'eau.

Malt (1994) a montré que les liquides sont appelés « eau » en fonction de leurs propriétés selon quatre dimensions, à savoir selon leur composition chimique, leur origine, leur emplacement, et leur utilisation. Elle a aussi montré que l'eau du robinet et l'eau des océans étaient des instances typiques d'eau (expérience 2). Elle conclut qu'un liquide est appelé « eau » et est jugé être de l'eau s'il est suffisamment semblable à l'eau du robinet ou à l'eau des océans du point de vue de sa composition chimique, de son origine, de son emplacement et de son utilisation. Malt écrit ainsi (Malt manuscrit, 12<sup>29</sup>) :

« A liquid that is pure H<sub>2</sub>O may be called 'water' because of the shared major component of chemical composition as well as its drinkability. A liquid that has less H<sub>2</sub>O and more organic matter but is in a lake might be called 'water' because of its overlap on source and partial overlap on uses and composition, though having a difference in current location. (Historically, in fact, it was probably ocean, lake, and river waters that provided the prototype, because these would have been encountered and named long before water came out of taps or in bottles. Over time, the prototype apparently has shifted for people living in the industrialized United States.) »

Ces résultats suggèrent que EAU<sub>NE</sub> est un prototype ou un ensemble d'exemplaires, ou encore qu'il y a plusieurs concepts non essentialistes, un prototype et un ensemble

---

<sup>29</sup> Malt, B. C., « Reflections on water (and « water ») », manuscrit.

d'exemplaires. En effet, les résultats de Malt montrent que la catégorisation des liquides comme de l'eau est semblable à la catégorisation fondée sur des prototypes (voir le chapitre 5, schéma 5.6.) ou sur des exemplaires (Medin et Schaffer 1978, Nosofsky 1986 – voir le chapitre 5, schéma 5.7.). Certains items sont typiques et les autres items sont jugés être de l'eau en fonction de leur ressemblance avec ces items.

En s'appuyant sur les résultats de Malt, on peut spéculer sur le prototype EAU<sub>Pr</sub>. J'utilise la théorie des prototypes proposée par Smith et al. 1988 (voir le chapitre 3, schéma 3.1.)<sup>30</sup> :

<b>Dimensions</b>	<b>Propriétés</b>
Structure chimique 0.6	H <sub>2</sub> O 10
	Minéraux 1
Origine 0.6	Source 8
	Ciel 6
	Océan 6
Emplacement 0.6	Bouteille 8
	Lac 8
	Océan 8
Utilisation 0.6	Boisson 10
	Nettoyage 4
Couleur 0.4	Transparent 10
	Marron 4
Odeur 0.4	Inodore 8
	Puant 2
Goût 0.4	Sans goût 10

**Schéma 7.3.**

<sup>30</sup> Cette description ne prétend pas décrire exactement notre prototype. Il s'agit plutôt d'une illustration.

Cette description résume les dimensions et les propriétés qui, selon Malt 1994, sont importantes dans notre catégorisation. Un liquide est jugé être de l'eau si sa ressemblance avec ce prototype est supérieur à un seuil de catégorisation. Bien sûr, les pondérations des dimensions et des traits proposées ci-dessus sont imaginaires.

Les résultats de Malt sont aussi compatibles avec l'idée suivante. EAU<sub>NE</sub> est en fait un ensemble d'exemplaires. Par exemple, EAU<sub>NE</sub> pourrait être la paire suivante d'exemplaires.

<b>EAU DU ROBINET</b>	
Composition	H <sub>2</sub> O Minéraux
Origine	Robinet
Emplacement	Robinet
Utilisation	Boisson Cuisine Nettoyage
Couleur	Transparent
Goût	Sans goût
Odeur	Chimique

**Schéma 7.4.**

<b>EAU DES OCEANS</b>	
Composition	H <sub>2</sub> O Sel
Origine	Océan
Emplacement	Océan

Utilisation	Baignade
Couleur	Bleu
Goût	Salé
Odeur	Marin

**Schéma 7.5.**

Un liquide est jugé être de l'eau s'il est suffisamment semblable à ces exemplaires.

On pourrait finalement aussi proposer que nous avons ces deux types de concepts non essentialistes. Cependant, je n'ai aucune donnée empirique à avancer en faveur de cette thèse. Elle est compatible avec les données empiriques de Malt.

Je conclus que EAU<sub>NE</sub> n'est pas un concept théorique d'artéfacts, comme Bloom l'avait proposé. Il ressemble davantage à un prototype ou à un ensemble d'exemplaires. Bien que Bloom souligne correctement que nous avons (au moins) deux concepts d'eau, sa description du concept non essentialiste est sans doute fausse.

### *3.4. L'hypothèse de l'hétérogénéité prend-elle l'eau ?*

J'ai montré que nous avons au moins deux concepts d'eau, un concept essentialiste d'eau, que j'ai appelé EAU<sub>E</sub> et au moins un concept non essentialiste d'eau, que j'ai appelé EAU<sub>NE</sub>. Ce concept n'est pas théorique : notre catégorisation de liquides comme étant de l'eau montre que c'est un prototype ou un ensemble d'exemplaires.

Je souhaite montrer maintenant que cette situation vient à l'appui de la prémisse 2 de l'ANE. L'argument est le suivant. Les concepts essentialistes sont des concepts théoriques. Ils attribuent des essences aux membres des catégories représentées. Cette information n'est encodée ni par les prototypes ni par les ensembles d'exemplaires. Elle est semblable aux informations théoriques. En outre, le contenu des concepts essentialistes est déterminé par des théories d'arrière-plan (voir le chapitre 3 sur cette notion) : nous attribuons différents types d'essences à des matières (composition matérielle)<sup>31</sup>, à des fruits (être le produit d'un arbre particulier)<sup>32</sup>, et à des animaux (avoir un type particulier d'ADN, être engendré par des animaux qui appartiennent à la même espèce, engendrer des animaux qui appartiennent à la

---

<sup>31</sup> Keil 1989.

<sup>32</sup> Keil 1989.

même espèce...)<sup>33</sup> et à des artefacts (être produit pour une certaine fin)<sup>34</sup>. Cela semble être le résultat de nos théories d'arrière-plan sur les domaines concernés, par exemple de notre théorie d'arrière-plan sur les artefacts, sur les végétaux, sur les animaux (biologie naïve).<sup>35</sup> Par ailleurs, j'ai admis que EAU<sub>NE</sub> était un prototype, un ensemble d'exemplaires ou les deux. En conséquence, la dissociation entre notre croyance que l'eau est par essence H<sub>2</sub>O et certaines pratiques de dénomination, ainsi que les divergences entre nos intuitions linguistiques sur les conditions de vérité des énoncés contenant le mot « eau », viennent à l'appui de la prémisse 2 de l'ANE.

### 3.5. Généralisation

La discussion dans ce chapitre s'est concentrée sur les concepts d'eau. Il faut souligner que les arguments avancés ici ne reposent nullement sur les propriétés particulières de nos concepts d'eau. La plupart des mots expriment plusieurs concepts qui appartiennent à des types fort différents de concepts (Quine 1977, Dupré 1981, 1986, 1999).

D'autres mots pourraient en effet illustrer la dissociation mise au jour par les travaux de Malt et les désaccords entre nos intuitions linguistiques. C'est, par exemple, le cas de « homme » :

- (10) Les castrats ne sont pas des hommes.
- (11) Les castrats sont des hommes.

Des expressions comme « techniquement parlant », « en un sens », « en un autre sens » etc. (appelées « hedges » en anglais) peuvent être utilisées pour produire ce type de contradiction pour de très nombreux mots. C'est le cas par exemple de « vie » :

- (12) Techniquement parlant, les zombies ne sont pas en vie.
- (13) Les zombies sont en un sens en vie.

Ce fait suggère que ces expressions sont associées à plusieurs concepts, comme cela est prédit par l'hypothèse de l'hétérogénéité.

---

<sup>33</sup> Keil 1989, Gelman et Wellman 1988, Hatano et Inagaki 1999, Au et Romo 1999.

<sup>34</sup> Bloom 1996.

<sup>35</sup> Atran 1990, Medin et Atran 1999.

### 3.6. Conclusion

Dans cette section, j'ai examiné différentes manières de décrire les concepts essentialistes et non essentialistes d'eau que les travaux de Malt nous ont conduit à postuler. Bloom a proposé que ces deux concepts étaient de nature théorique et résultaient de deux manières d'appréhender la réalité, en termes d'espèces naturelles ou en termes d'artéfacts. Gelman et Hirschfeld proposent que l'un de ces concepts est générique, tandis que l'autre est spécifique. J'ai cherché à montrer que le concept non essentialiste est en fait un prototype ou un ensemble d'exemplaires, peut-être les deux. Le concept essentialiste est un concept théorique. En conséquence, la prémisse 2 de l'ANE s'accorde avec les travaux de Malt et avec les contradictions entre nos intuitions linguistiques.

## 4. Conclusion

Résumons cette longue discussion des travaux de Malt. Le but de ce chapitre était de défendre une prédiction de l'hypothèse de l'hétérogénéité et de la prémisse 2 de l'ANE : il s'agissait de montrer que la plupart des expressions linguistiques expriment plusieurs concepts, entre autres un prototype, une théorie et un ensemble d'exemplaires. Je me suis servi de « eau » comme exemple, parce que les travaux de Malt ont été largement discutés.

Nous sommes des essentialistes, comme le soutient la théorie de l'essentialisme psychologique. Nous supposons ainsi que l'eau possède une essence, H<sub>2</sub>O. Cette croyance joue un rôle dans de nombreux contextes, comme le montrent les travaux de Keil et nos intuitions linguistiques. C'est donc un concept (voir le chapitre 1, section 1.2.3.).

Malt a par ailleurs montré que certaines pratiques de dénomination et certains jugements de catégorisation ne prenaient pas en compte ce concept essentialiste. Bien que nous croyions que l'eau soit H<sub>2</sub>O, nous n'utilisons pas cette croyance pour juger si le thé, le sang, la sueur etc. sont de l'eau. Ces données ne peuvent être expliquées en faisant appel aux règles pragmatiques de communication. Il faut donc postuler l'existence d'un concept non essentialiste d'eau.

J'ai finalement montré que le concept essentialiste était de nature théorique, tandis que le concept non essentialiste était un prototype ou un ensemble d'exemplaires. En conséquence, « eau » exprime au moins deux concepts différents, un concept théorique et essentialiste, utilisé par exemple dans les raisonnements à voix haute, et un concept qui est un prototype ou un ensemble d'exemplaires. Même si ces résultats n'établissent pas sans

ambiguïté que nous avons des théories, des prototypes *et* des exemplaires, ils sont compatibles avec la prédiction dérivée de la prémisse 2 de l'ANE.

## Chapitre 8

### Contre les théories composites des concepts

---

L'argument de la nécessité explicative repose sur la thèse suivante : la *meilleure* explication des propriétés des processus cognitifs supérieurs humains postule plusieurs types de représentations mentales qui ont peu en commun – en particulier des prototypes, des exemplaires et des théories. Dans les chapitres 5 et 6, j'ai cherché à montrer qu'il est possible d'expliquer de cette manière les phénomènes mis au jour par l'étude empirique de la catégorisation et de la composition des concepts. J'ai aussi montré que ces phénomènes ne pouvaient être expliqués par une théorie qui ne poserait que des prototypes, que des exemplaires ou que des théories. Cependant, j'ai aussi reconnu qu'une théorie composite des concepts (TCC) était aussi en mesure d'expliquer ces données.<sup>1</sup> Les théories composites des concepts soutiennent que les concepts encodent différents types de connaissances. Considérons, en particulier, la théorie composite suivante : un concept encode des connaissances à propos des propriétés typiques des membres de la catégorie dénotée, des connaissances à propos de certains membres de cette catégorie et des connaissances théoriques à propos de la catégorie. De prime abord, une telle théorie explique aussi bien les données sur la composition des concepts et sur la catégorisation que l'hypothèse de l'hétérogénéité (HH). Car l'HH et cette théorie composite des concepts supposent que les concepts encodent les mêmes connaissances.<sup>2</sup> On peut en conséquence douter que l'HH soit la meilleure explication des propriétés des processus cognitifs supérieurs.

Bien plus, des considérations de parcimonie semblent favoriser cette théorie composite. Cette idée a été présentée au chapitre 5 (chapitre 5, section 5.2.) :

#### **L'argument de la parcimonie pour les théories composites des concepts**

1. Certaines théories composites des concepts supposent qu'un concept encode exactement les connaissances qu'un prototype, qu'un ensemble d'exemplaires et qu'une théorie coréférentiels encoderaient.

---

<sup>1</sup> Voir le chapitre 5, section 5.

<sup>2</sup> Dans le reste de ce chapitre, « TCC » et l'expression au singulier « la théorie composite des concepts » désignent cette théorie composite particulière. L'expression au pluriel « les théories composites des concepts » désigne la classe entière des théories composites des concepts.

2. L'étude d'une compétence cognitive supérieure, par exemple la catégorisation, montre que cette compétence repose sur les types de connaissances que les prototypes, les exemplaires et les théories sont supposés encoder.
3. La prémisse 2 vient également à l'appui de l'HH et de la théorie composite des concepts (d'après 1).
4. La théorie composite des concepts est plus parcimonieuse.
5. Donc, on devrait admettre cette théorie (d'après 3 et 4).

En conséquence, la prémisse 2 de l'argument de la nécessité empirique (il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs) ne sera vraiment établie que s'il est possible de rejeter cette TCC. C'est l'objet de ce chapitre. Le cœur de l'argument est le suivant. Les avocats de la TCC doivent expliquer ce que signifie la thèse selon laquelle différents ensembles de connaissances sont *des parties* d'un même concept, et non des concepts différents. Cette thèse revient de prime abord à soutenir que les différents ensembles de connaissance sont fonctionnellement hiérarchisés : l'un d'entre eux définit les conditions d'application des autres ensembles de connaissances. Si les avocats de la TCC admettent cette interprétation, je soutiens que peu de données empiriques viennent à l'appui d'une telle affirmation. S'ils rejettent cette interprétation, je soutiens que la théorie composite des concepts est seulement une variante terminologique de l'HH. Dans ce cas, les avocats de la TCC rejettent de fait l'hypothèse EN.

Dans la première section, je présente la notion d'indépendance fonctionnelle et l'argument qui repose sur cette notion. Dans la seconde section, je montre que peu de données empiriques viennent à l'appui de la thèse selon laquelle les différents ensembles de connaissances sont fonctionnellement hiérarchisés. Dans la dernière section, je montre que si les avocats de la TCC rejettent la notion d'indépendance fonctionnelle, la TCC et l'hypothèse de l'hétérogénéité ne sont que des variantes terminologiques.

## **1. Le dilemme de l'indépendance fonctionnelle**

Dans cette section, j'explicité la notion d'indépendance fonctionnelle. Sur cette base, je présente le dilemme dans lequel la théorie composite des concepts est empêtrée.

### 1.1. L'indépendance fonctionnelle

Supposons qu'un agent ait deux ensembles de connaissances à propos des chiens, A et B, par exemple des connaissances à propos de certains chiens particuliers et des connaissances à propos des propriétés essentielles des chiens. A et B sont fonctionnellement indépendants si et seulement si deux conditions sont satisfaites :

- (i) On accomplit différemment une tâche cognitive, par exemple on catégorise différemment une cible dans la catégorie des chiens, quand on fait appel à A et à B.
- (ii) Il n'y a pas de hiérarchie entre A et B.

Explicitons ces deux parties de la définition. Que veut dire (i) ? (i) revient à dire que A et B sont deux ensembles de connaissances différents. En conséquence, les processus cognitifs dans lesquels ils figurent se déroulent différemment. Car le déroulement d'un processus cognitif dépend des connaissances qui sont extraites de la MLT (voir le chapitre 1, section 2.3. et schéma 1.3.).

Comment comprendre (ii) ? A et B spécifient deux manières différentes d'accomplir une tâche cognitive, par exemple catégoriser une cible dans la catégorie des chiens (voir (i)). L'idée est que ni A ni B ne sont traités par l'esprit comme spécifiant la manière correcte de remplir les tâches cognitives. Si faire appel à A et faire appel à B mènent à des résultats différents dans une tâche cognitive donnée, l'esprit ne traite ni A ni B comme fournissant *la* manière correcte d'accomplir cette tâche. Supposons par exemple que l'agent ait à catégoriser une cible sur la base d'exemplaires de chiens (A) ou sur la base des propriétés des chiens tenues pour essentielles (B), par exemple le fait que les chiens engendrent des chiens.<sup>3</sup> La cible est un animal qui a été engendré par des chiens, mais qui ne ressemble pas aux chiens représentés par les exemplaires stockés dans la mémoire à long terme de l'agent. Selon B, c'est un chien. Selon A, ce n'est pas un chien. (ii) signifie que l'agent ne tient ni A ni B pour la manière correcte de catégoriser la cible. Dans ces circonstances, je dis que ces deux ensembles de connaissances sont fonctionnellement indépendants.

---

<sup>3</sup> Gil-White écrit à propos de nos concepts théoriques d'espèces animales (Gil-White 2001, 522 – voir le chapitre 7, section 1.) : « In species categories, the possession of an essence of X, I argue, promotes a strongly held intuition that one mates with X and produces Xs in reproduction. Thus, whether a token has an essence of X answers to questions (1) Does it produce Xs in reproduction ? (2) Is it descended from Xs ? and (3) Does it mate with Xs ? »

Pour bien saisir cette idée, considérons une position qui rejetterait l'idée que des ensembles de connaissances à propos d'une même catégorie sont fonctionnellement indépendants. C'est le cas des théories duelles des concepts (voir le chapitre 2, section 3. et la section 2.1. ci-dessous). Selon ces théories, nos concepts sont composés de deux parties, une définition et un prototype. Le prototype sert de procédure de reconnaissance utilisée dans les catégorisations instantanées : c'est une manière approximative, souvent efficace, d'identifier les instances du concept. Quand un agent catégorise une cible dans la catégorie des chiens sur la base de son prototype de chien, son jugement est provisoire et peut être corrigé (« defeasible »). L'agent lui-même considère que le cœur du concept, à savoir la définition, fournit le critère définitif à propos de l'appartenance de la cible à la catégorie des chiens. Si les deux ensembles de connaissances produisent des catégorisations différentes, l'agent suppose que la catégorisation fondée sur la définition est la bonne. Dans ce cas, je dis que les deux ensembles de connaissances sont hiérarchiquement structurés et ne sont pas fonctionnellement indépendants.

Résumons. La notion d'indépendance fonctionnelle peut être définie comme suit :

- (IF) Deux ensembles de connaissances A et B sont fonctionnellement indépendants si et seulement si l'agent ne traite ni l'un ni l'autre comme fournissant à lui seul la manière correcte d'accomplir les tâches cognitives.

### 1.2. Le dilemme

L'indépendance fonctionnelle est au cœur de l'argument que j'oppose à la théorie composite des concepts. La TCC doit expliquer ce que signifie l'idée que les ensembles de connaissances qu'elle postule sont des *parties* d'un *même* concept plutôt que *des concepts différents*.

- (a) Si cela signifie que ces ensembles de connaissances ne sont pas fonctionnellement indépendants, c'est-à-dire sont hiérarchisés, alors peu de données empiriques viennent à l'appui de la TCC.
- (b) Si la TCC admet que ces ensembles de connaissances sont fonctionnellement indépendants, elle ne diffère de l'hypothèse de l'hétérogénéité que terminologiquement.

En bref, ou bien l'HH et la TCC sont simplement des variantes terminologiques ou bien peu de données empiriques viennent à l'appui de la TCC. Le reste de ce chapitre a pour objet de défendre les deux branches de ce dilemme.

### *1.3. Conclusion*

Dans la première section, j'ai présenté la notion d'indépendance fonctionnelle et le dilemme contre la théorie composite des concepts. Deux ensembles de connaissances à propos d'une même catégorie sont fonctionnellement indépendants si et seulement si ni l'un ni l'autre ne sont traités par l'agent comme étant la manière correcte d'accomplir les tâches cognitives. Si la TCC reconnaît que les parties des concepts sont fonctionnellement indépendantes, je soutiens que la différence entre la TCC et l'HH est purement terminologique. Si elle soutient que les parties des concepts sont hiérarchisées, je soutiens que peu de données empiriques viennent à son appui. L'HH est alors supérieure à la TCC.

## **2. Théories composites et indépendance fonctionnelle**

Dans cette section, je montre que les avocats des théories composites supposent que les ensembles de connaissances censés former les différentes parties des concepts sont hiérarchisés. Bien que l'objet de ce chapitre soit une théorie composite particulière, à savoir la TCC, je passe en revue différentes théories composites. Je montre par là que la hiérarchisation des parties motive en général l'adhésion aux théories composites des concepts : soutenir une théorie composite des concepts, c'est en fin de compte embrasser la hiérarchisation des différents ensembles de connaissances relatifs à une même catégorie.

### *2.1. Les théories duelles des concepts*

#### 2.1.1. Que soutiennent les théories duelles des concepts ?

L'idée de base des théories duelles des concepts est bien exprimée par la citation suivante de deux des principaux avocats des théories duelles (Osherson et Smith 1981, 277-278, voir le chapitre 2, section 3. et ci-dessus).<sup>4</sup>

« (...) prototype theory is by its nature incomplete because it is about only a limited aspect of concepts. To make this clearer, we can distinguish between a concept's *core* and its *identification procedure* ; the core is concerned with those aspects of a concept

---

<sup>4</sup> Voir aussi Miller et Johnson-Laird 1976.

that explicate its relation to other concepts, and to thoughts, while the identification procedure specifies the kind of information used to make rapid decisions about membership. (This distinction is similar to the one proposed by Miller and Johnson-Laird, 1976). We can illustrate with the concept *woman*. Its core might contain information about the presence of a reproductive system, while its identification procedures might contain information about body shape, hair length, and voice pitch. Given this distinction it is possible that some traditional theory of concepts correctly characterizes the core, whereas prototype theory characterizes an important identification procedure. »

Nous pouvons donc résumer les théories duelles comme suit (voir aussi Malt et Johnson 1992, 196).

### **Théories duelles**

1. Les concepts ont deux parties, un cœur et une procédure d'identification.
2. Les exemples utilisés par les avocats des théories duelles suggèrent que le cœur est une définition ou un ensemble de propriétés tenues pour nécessaires.
3. La procédure d'identification est un prototype ou un ensemble d'exemplaires.
4. Les deux parties sont hiérarchiquement organisées : le cœur est traité par le possesseur du concept comme un critère de correction.

Dans le présent contexte, 4 est bien sûr l'idée essentielle. Quand un agent catégorise une cible comme une grand-mère, il peut utiliser son prototype de grand-mère : il peut par exemple catégoriser une vieille femme aux cheveux gris comme une grand-mère, parce qu'elle est semblable au prototype de grand-mère. Mais selon les théories duelles des concepts, l'agent considère que la définition « être la mère d'un parent », c'est-à-dire le cœur du concept, fournit le critère définitif d'appartenance à la catégorie des grand-mères. Si l'agent découvrait que la cible qui ressemble au prototype de grand-mère n'est en fait pas la mère d'un parent, il retirerait son jugement de catégorisation.<sup>5</sup> Les deux parties sont donc hiérarchisées.

Cette idée est clairement exprimée par Armstrong, Gleitman et Gleitman (1983, 248) :

---

<sup>5</sup> Cela semble intuitivement plausible. Voir cependant ci-dessous.

« Perhaps the graded judgments and responses have to do with a mentally stored *identification function* used to make quick sorts of things, scenes, and events in the world. On this formulation, instances of a concept share some rough and ready list of perceptual and functional properties, to varying degrees (...). For example, grandmothers tend to have *grey hair, a twinkle in their eye*. Some of these properties may be only loosely, if at all, tied to the criterion for membership in the class (for example, *twinkles* for grandmotherliness) while others may be tightly, systematically, tied to the criteria for membership (for example, being *adult* for grandmotherliness). But in addition to this identification function, there will be a mentally stored *categorical description* of the category that does determine membership in it. For *grandmother*, this will be *mother of a parent*. »

### 2.1.2. La critique standard

Ces théories ont à juste titre fait long feu. Dans la mesure où les théories duelles supposent que le cœur est une définition, elles héritent de toutes les difficultés des théories classiques des concepts (Laurence et Margolis 1999) : nous sommes incapables de définir les catégories (Fodor et al. 1980) et, ce qui est plus important, nous n'utilisons que très rarement des définitions dans nos processus cognitifs (Murphy 2002).

### 2.1.3. Limites des théories duelles

Les théories duelles des concepts pourraient postuler que le cœur des concepts encode des propriétés tenues pour nécessaires plutôt que des définitions complètes. Par exemple, le cœur de CHIEN pourrait spécifier des propriétés des chiens tenues pour nécessaires, par exemple *être engendré par des chiens*, plutôt qu'une définition. Il est plus difficile de rejeter ce type de théorie duelle : j'ai en effet montré au chapitre précédent que dans certaines tâches cognitives, par exemple quand nous raisonnons à voix haute (ainsi que dans les expériences de Keil), nous utilisons bien des propriétés nécessaires.

Ces théories duelles sont cependant problématiques, car elles sont incapables d'expliquer les données empiriques présentées dans les chapitres précédents. Ces données montrent (i) que nous encodons des connaissances à propos de certains membres des catégories, des connaissances à propos des propriétés typiques et des connaissances théoriques et (ii) que nous utilisons régulièrement ces connaissances dans nos processus supérieurs. Les théories duelles ne sont pas supposées encoder ce type de connaissances : leur pouvoir explicatif est donc limité.

## 2.2. Les théories composites contemporaines

La psychologue développementale Rochel Gelman a présenté une théorie composite des concepts selon laquelle les concepts encodent précisément les trois types de connaissances auxquels nos processus cognitifs font par défaut appel (Gelman 1990, 2001, voir aussi Schwartz et Reisberg 1991, cité dans Gelman 2001) :

« The proposal that a mental learning tool computes the contingent frequency of relevant encounters fits with conclusions drawn by other authors. Schwartz and Reisberg (1991, p. 391) suggested that researchers may need a three-part theory of concepts, in which ‘concepts are represented by a prototype, some set of specifically remembered cases, and some further abstraction,’ where the parts all interact to accomplish correct similarity judgments and inferences. » (Gelman 2001, 20)

Cette théorie est naturellement la principale alternative de l’HH. Je l’ai appelée plus haut « la TCC ».

La théorie proposée par Gelman peut être résumée comme suit.

### **La théorie composite de Gelman (TCC)**

1. Les concepts sont composés de trois parties : une partie théorique, une partie prototypique et une représentation de certaines instances du concept.
2. La partie théorique peut spécifier des définitions ou des propriétés nécessaires ; elle peut aussi exprimer des propositions causales ou fonctionnelles.
3. La partie théorique spécifie quelles connaissances sont encodées par les deux autres parties.<sup>6</sup>
4. Les trois parties sont hiérarchisées : la partie théorique sert de critère.

Dans le présent contexte, 4 est l’idée importante. Gelman insiste précisément sur cette idée. Elle écrit (Gelman 2001, 19-20) :

---

<sup>6</sup> Parce qu’elle détermine à quelles propriétés nous prêtons attention. Voir ci-dessous, section 3.2..

« Should a huge, hard-surfaced mass that is in a garden start to move, the observer will undoubtedly no longer think it is a rock. Rocks simply cannot cause themselves to move – although turtles can, even if they are in an unexpected setting. Therefore, statistical information, by itself, does not serve to define concepts. »

### 2.3. Conclusion

Dans cette section, j'ai présenté les deux principales théories composites des concepts, à savoir les théories duelles et la théorie composite de Gelman. Ces deux théories admettent l'idée d'une hiérarchisation des parties des concepts : les prototypes et/ou les exemplaires correspondent à des parties qui sont fonctionnellement dépendantes de l'autre partie – une définition ou une théorie. Je propose que cette idée motive *en général* l'adhésion aux théories composites des concepts : l'une des parties des concepts spécifie les conditions correctes d'utilisation des autres parties.

#### **L'objet des théories composites des concepts**

Soutenir que nos ensembles de connaissances à propos d'une même catégorie sont *des parties d'un même concept* plutôt que *des concepts différents*, c'est affirmer qu'ils sont *hiérarchisés*.

## 3. Quelle est la validité empirique de la TCC ?

Dans cette section, je présente et critique les données qui semblent venir à l'appui de la TCC. Je montre qu'il n'y a aucune raison de préférer la TCC à l'hypothèse de l'hétérogénéité.

### 3.1. Ressemblance et appartenance : une distinction intuitive

#### 3.1.1. L'argument

De nombreux lecteurs ont probablement l'intuition suivante : être la mère d'un parent est le critère de correction des décisions de catégorisation dans la catégorie des grands-mères (Armstrong et al. 1983, Gelman 2001). Ces lecteurs supposent sans doute également que cette intuition est partagée par tout un chacun. Le philosophe G. Rey a insisté sur ces intuitions (1983, 287, 1985, voir aussi Gelman et Markman 1987, Gil-White 2001) :

« (...) what is really at issue is whether there is any reason to distinguish the metaphysical from the epistemological roles concepts might play. (...) There is first of

all the point that it does appear to be a *psychologically real* distinction : people do distinguish [looking like a bird] from [actually being a bird]. At any rate, as we have already noted, most adults know that something might exhibit the outward, ‘perceptible’ and/or ‘accidental’ features of a bird, but be in fact a mechanical toy and so not any sort of bird at all. And conversely (...) people do count things that *do not* look like birds – e.g., penguins – as nevertheless real birds. »

Puisque nous distinguons entre l’apparence d’un objet et l’appartenance réelle à une catégorie, Rey conclut que les concepts ne peuvent être ni des prototypes ni des exemplaires.<sup>7</sup> Adaptons cet argument au présent contexte :

### **L’argument apparence/appartenance**

1. Nous utilisons des prototypes et des exemplaires pour catégoriser.
2. Nous ne traitons pas la ressemblance avec un prototype ou avec des exemplaires comme le critère d’appartenance à une catégorie : nous croyons qu’il existe des conditions d’appartenance différentes – même si nous pouvons concéder ignorer leur nature.
3. Donc, dans les tâches de catégorisation, les ensembles de connaissances à propos des propriétés typiques et les connaissances à propos de membres particuliers de la catégorie dépendent d’autres connaissances.
4. Ces connaissances sont des définitions ou font référence à la possession d’une essence.
5. Donc, la TCC décrit correctement nos concepts.

#### 3.1.2. Réponse

Comme tout un chacun, je partage l’intuition sur laquelle cet argument repose. J’ai aussi l’intuition qu’une vieille femme aux cheveux gris n’est pas une grand-mère si elle n’est pas la mère d’un parent et que Tina Turner est une grand-mère même si ses cheveux ne sont pas gris. Pareillement, il y a des oiseaux atypiques. J’accepte donc la prémisse 2 de l’argument présenté ci-dessus. Mais je doute que ces intuitions montrent en général que les différents ensembles de connaissances à propos d’une même catégorie soient hiérarchisés. Ainsi, je rejette l’inférence de la thèse 3 à partir de la prémisse 2.

---

<sup>7</sup> Rey suppose à tort que les théories des prototypes sont nécessairement empiristes (voir aussi Gelman et Markman 1987, 153).

Il faut tout d'abord concéder que nos différents ensembles de connaissances à propos de *certaines* catégories sont en effet hiérarchisés. En d'autres termes, pour ces catégories, la TCC est sans doute correcte. C'est peut-être le cas des catégories dont nous connaissons certaines conditions nécessaires d'appartenance. Par exemple, les euros français doivent être produits par la banque de France ; autrement, ce sont des faux billets.

Mais cela n'est le cas que d'un petit nombre de catégories. Pour les autres catégories, la prémisse 2 n'est que l'expression de notre essentialisme naïf et de notre possession de concepts essentialistes. Notre concept essentialiste de grand-mère exprime la proposition qu'une grand-mère est nécessairement la mère d'un parent (Armstrong et al. 1983). Notre concept théorique de rocher spécifie que les pierres ne peuvent se mouvoir elles-mêmes (Gelman 2001). Notre concept théorique d'oiseau exprime sans doute la proposition que les oiseaux sont en vie et ont été engendrés par d'autres oiseaux (Rey 1983). Notre concept essentialiste d'eau spécifie que l'eau est composée de molécules dont la structure chimique est H<sub>2</sub>O (voir le chapitre 7). Quand nous extrayons notre concept théorique GRAND-MERE etc. de notre MLT, nous faisons précisément la distinction entre une grand-mère réelle et une grand-mère apparente. Parce que nous avons ces concepts essentialistes ou théoriques, nous avons donc les intuitions auxquelles la prémisse 2 fait référence : Tina Turner est une grand-mère et les larmes sont de l'eau.

Le fait que nous ayons des intuitions essentialistes ne montre toutefois pas que nos concepts essentialistes sont utilisés comme des critères de correction de nos catégorisations, raisonnements etc. Comme nous l'avons vu, les études de Malt montrent au contraire que notre concept non essentialiste d'eau (ou nos concepts non essentialistes d'eau – voir le chapitre 7), à savoir un prototype d'eau et/ou un ensemble d'exemplaires d'eaux, ne sont pas fonctionnellement dépendants de notre concept essentialiste d'eau (voir le chapitre 7). Malt et Johnson (1992), Smith et Sloman (1994) et Hampton (1995) montrent pareillement que dans de nombreux contextes, nos catégorisations ne dépendent pas de nos concepts essentialistes.<sup>8</sup> Nous croyons que nous catégorisons un liquide comme étant de l'eau si nous savons qu'il est presque exclusivement composé de molécules d'H<sub>2</sub>O, mais nous ne prenons pas en compte

---

<sup>8</sup> Malt et Johnson écrivent (1992, 210) : « The results of these experiments suggest that having a particular function does not constitute either a necessary or a sufficient condition for membership in an artifact category, and that function therefore does not provide a core for artifact concepts in the strongest sense. The results also suggest that function may not provide a core in the weaker sense of being more heavily weighted in category membership decisions than other kinds of information. »

cette croyance quand nous devons effectivement catégoriser. En conséquence, cette intuition ne montre pas que les différents ensembles de connaissances sont en fait hiérarchisés.

Deuxièmement, si nos concepts essentialistes étaient en effet tenus pour les critères de correction, on ne pourrait expliquer naturellement que nous tenions les deux énoncés suivants pour vrais (voir le chapitre 7, section 3.5. pour d'autres exemples) :

1. Tina Turner est en un sens une grand-mère.
2. Tina Turner n'est en un sens pas une grand-mère.

Ce type d'intuitions linguistiques vient fortement à l'appui de ma position. Le même mot est associé à différentes représentations, si bien que nous jugeons vrais des énoncés superficiellement contradictoires. Cela montre plutôt que nos concepts théoriques ne sont pas tenus pour les critères de nos jugements de catégorisation.<sup>9</sup>

L'argument pour la théorie composite des concepts présenté par Rey repose donc sur *une illusion cognitive*. Nous croyons spontanément que l'eau est essentiellement H<sub>2</sub>O etc. Parce que les psychologues et les philosophes ont les mêmes intuitions – ils ont en effet les mêmes concepts théoriques, ils supposent que ces concepts fournissent les critères pour évaluer nos performances.<sup>10</sup> Ces scientifiques sont trompés par leurs propres intuitions essentialistes : les études de Malt et autres montrent au contraire que nos pratiques de dénomination et nos jugements de catégorisation ne dépendent pas toujours de nos concepts essentialistes.<sup>11</sup>

Je conclus donc que l'intuition qui distingue entre une femme qui ressemble à une grand-mère et une grand-mère réelle ne montre pas que notre prototype de grand-mère ou notre ensemble d'exemplaires de grand-mères est fonctionnellement dépendant de notre concept essentialiste. Au contraire, certaines intuitions linguistiques montrent plutôt que les ensembles de connaissances à propos d'une même catégorie ne sont pas hiérarchisés.

---

<sup>9</sup> Ces intuitions peuvent être expliquées autrement. Cependant, ces autres explications me semblent plus compliquées que l'explication que je propose.

<sup>10</sup> Gil-White écrit ainsi (2001, 520) : « Critical to species categories, (...) are necessary and sufficient conditions of membership. (...) But knowing this does not mean that we always know an X when we see one – at least not at first. We may initially guess wrong but, upon learning relevant background facts, recognize our initial mistake and revise the initial classification. What this implies is that appearances may guide our fuzzy guesses when we try to sort new stimuli into categories but that these same appearances will – in the case of species categories – not impose themselves on our concept representation, which is unequivocal rather than fuzzy. »

<sup>11</sup> Ces idées suscitent souvent de fortes réticences. Elles résultent peut-être en partie de ce que les concepts théoriques sont fortement favorisés dans la culture occidentale (Malt 1993, voir aussi Nisbett 2003).

### 3.2. *L'argument de la coordination*

#### 3.2.1. L'argument

Supposons que vous voyiez un animal qui ressemble à un chien. Vous n'êtes pas un expert canin. Vous jugeriez sans doute que c'est un chien sur la base de sa ressemblance avec les chiens que vous avez rencontrés ou sur la base de la possession de certaines propriétés typiques des chiens. Vous pourriez peut-être alors conclure qu'il a été engendré par des chiens, qu'il possède une essence à l'intérieur de son corps, bien que vous soyez prêt à concéder que vous ne connaissez pas la nature de cette essence. Vous pourriez inférer qu'il possède tel et tel organe sur la base de votre connaissance générique des chiens etc. En d'autres termes, bien que vous ayez catégorisé cette cible sur la base de votre prototype ou sur la base d'un ensemble d'exemplaires de chiens, vous pourriez tirer des conclusions sur la base de vos connaissances théoriques. Le contraire est bien sûr également vrai. Supposons que vous catégorisiez une cible comme un chien parce que vous savez qu'elle a été engendrée par des chiens. Vous pourriez alors inférer qu'elle possède vraisemblablement les propriétés typiques des chiens ou qu'elle ressemble à Médor sous tel et tel aspect.

Ces exemples suggèrent que nos différents ensembles de connaissances sont *coordonnés*. J'utilise la notion de coordination comme suit :

Un ensemble de connaissances A est coordonné à un ensemble de connaissances B si et seulement si l'utilisation de A dans un processus cognitif dispose l'agent à extraire B de la MLT.

La relation de coordination n'est pas nécessairement symétrique. Et il peut y avoir différents types de coordination.

La coordination de nos ensembles de connaissances est un phénomène psychologique réel qu'il faut expliquer. On peut avoir l'impression qu'il supporte la TCC.

#### **L'argument de la coordination**

1. Il est fréquent que nos ensembles de connaissances à propos d'une même catégorie soient coordonnés.
2. Selon la TCC, ces ensembles de connaissances sont des parties hiérarchisées d'un même concept.

3. Ils sont donc coordonnés.
4. L'hypothèse de l'hétérogénéité n'implique pas la coordination de ces ensembles de connaissances.
5. Donc, la TCC fournit une meilleure explication du phénomène de coordination que l'hypothèse de l'hétérogénéité.

En d'autres termes, la TCC, mais pas l'HH, implique le phénomène psychologique de coordination. C'est donc une meilleure explication. Pourquoi la TCC implique-t-elle la coordination (prémises 2 et 3) ? Parce que les différentes parties des concepts sont supposées être fonctionnellement hiérarchisées. La hiérarchisation des parties des concepts suppose leur coordination. L'agent traite en effet l'une des parties, la partie théorique, comme fournissant le critère de l'accomplissement correct des tâches cognitives. De prime abord, la partie théorique ne peut avoir ce rôle que si les différentes parties des concepts sont coordonnées.

### 3.2.2. Réponse

Il est indéniable que nos différents ensembles de connaissances à propos d'une même catégorie sont souvent coordonnés. Les agents croient que le liquide qui est constitué de molécules dont la structure atomique est  $H_2O$  est typiquement transparent, inodore, sans goût, semblable à l'eau du robinet dans leur appartement etc. J'admets donc la prémisse 1 de l'argument. Cependant, je nie que cela vienne davantage à l'appui de la TCC que de l'HH. Les prémisses 2, 3 et 4 sont aussi vraies. La TCC, mais pas l'HH, implique bien la coordination de nos ensembles de connaissances.

Cependant, cet argument est problématique. La coordination entre différents ensembles de connaissances à propos d'une même catégorie ressemble en effet à la coordination entre nos concepts d'animal et nos concepts de chien, ou entre nos concepts de rottweiler et nos concepts de chiens. En d'autres termes, la coordination entre les ensembles de connaissances à propos des chiens ne se distingue pas de la coordination entre des concepts différents, CHIEN et ANIMAL par exemple. Or, par principe, l'hypothèse que les différents ensembles de connaissances à propos d'une même catégorie forment des parties hiérarchisées d'un même concept ne s'applique pas à la coordination entre différents concepts. Elle n'est donc pas à même de rendre compte de la coordination entre les différents ensembles de connaissances à propos d'une même catégorie. En outre, s'il est vrai que l'HH ne prédit pas cette coordination, elle n'est pas incompatible avec elle, puisque différents concepts peuvent être coordonnés.

Développons un peu cette réponse. Le premier point est que nos concepts sont massivement coordonnés. Si je sais que Médor est un rottweiler, je peux conclure que la plupart des choses que je connais à propos des chiens valent aussi pour Médor. Si je sais que Jerry est un crocodile, je peux conclure que la plupart des choses que je sais à propos des animaux valent aussi pour Jerry. Si je sais qu'Amélie Mauresmo est une joueuse de tennis, je peux conclure que la plupart des choses que je connais sur les athlètes, les êtres humains, les mammifères etc. valent aussi pour Mauresmo. Ainsi de nombreux concepts sont-ils coordonnés. Mais de toute évidence, les avocats de la TCC ne soutiennent pas que ce sont des parties d'un unique concept. Pourquoi donc la coordination de mes connaissances à propos de certains chiens, de mes connaissances des propriétés typiques des chiens et de mes connaissances théoriques à propos des chiens montre-t-elle que ces connaissances forment des parties d'un seul concept ?

Pour qui adopte l'HH, la coordination entre nos différents ensembles de connaissances à propos de la même catégorie, par exemple à propos des chiens, est seulement un cas particulier du fait général que les représentations mentales tendent à être coordonnées. Les agents utilisent de manière efficace leurs connaissances et cherchent à ne pas négliger des connaissances qui pourraient être utiles pour accomplir une tâche cognitive.

En réponse, un avocat de la théorie composite des concepts insisterait sans doute sur le fait que la coordination entre les différents ensembles de connaissances à propos d'une même catégorie est différente de la coordination entre différents concepts comme ANIMAL et CHIEN. Car l'agent sait que ces ensembles de connaissances sont à propos de la même catégorie. Il sait que la définition de grand-mère, le prototype de grand-mère et les représentations singulières de certaines grand-mères sont à propos de la même catégorie d'individus. Cela n'est bien sûr pas le cas d'un concept de chiens et d'un concept de rottweiler.

Certes, mais est-ce que cela montre que ces ensembles de connaissances à propos d'une même catégorie sont des parties d'un même concept ? La réponse est négative. Tout le monde reconnaît qu'un agent peut avoir deux concepts coréférentiels. Pour reprendre les exemples de Frege, « étoile du matin » et « étoile du soir » sont censés exprimer deux concepts différents de Vénus. Or, un agent qui sait que ces deux expressions dénotent la même planète cesse-t-il d'avoir deux concepts ? Pas nécessairement.<sup>12</sup> Si l'avocat de la TCC

---

<sup>12</sup> Quoique cela puisse arriver (Millikan 2000).

admet ce point, il n'a aucune raison de nier que nos ensembles de connaissances sur une même catégorie constituent différents concepts.

### 3.2.3. Coordination de nos concepts coréférentiels

Cette discussion soulève le problème suivant : si nos ensembles de connaissances à propos d'une même catégorie sont bien des concepts, comment ces concepts sont-ils coordonnés ? La question est bien sûr empirique. Si l'hypothèse de l'hétérogénéité est confirmée, cette question devrait être étudiée en détail par les psychologues. Etant donné la prévalence de l'hypothèse EN, quasiment aucune donnée empirique n'est disponible sur cette question. On peut néanmoins faire les suggestions suivantes.

- On peut supposer que le prototype, l'ensemble d'exemplaires et la théorie à propos d'une même catégorie sont *directement* coordonnés : l'extraction de l'un de ces ensembles de connaissances rend probable l'extraction des autres ensembles de connaissances sans que cette relation ait pour intermédiaire l'extraction d'une autre représentation. Le schéma suivant résume cette idée.

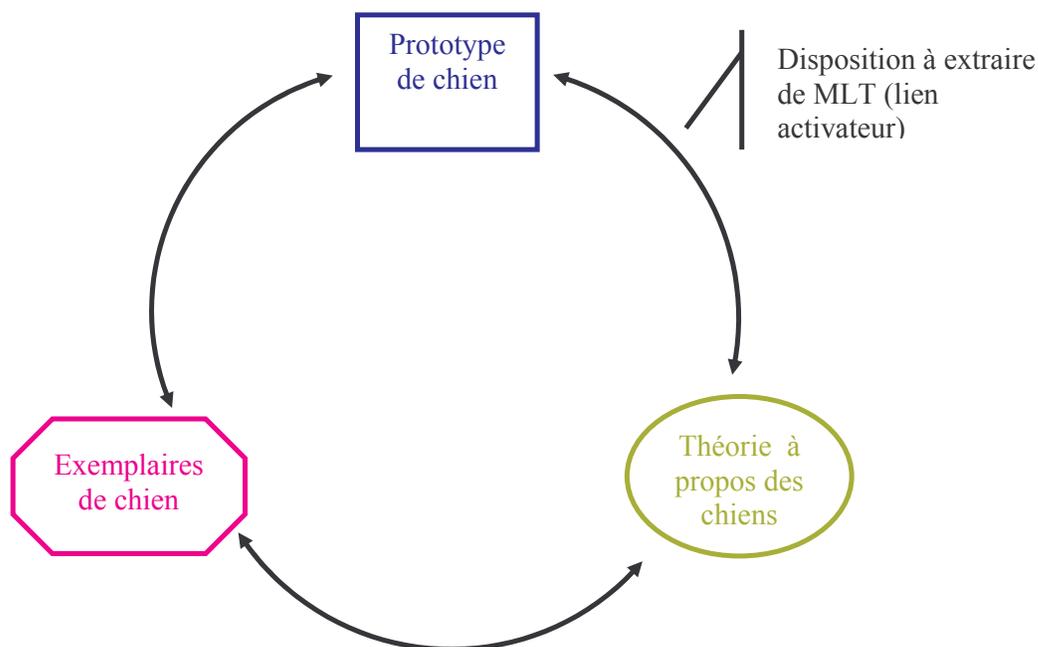


Schéma 8.1.

- Alternativement, on peut supposer que le prototype, l'ensemble d'exemplaires et la théorie d'une même catégorie sont *indirectement* coordonnés : l'extraction de l'un de ces ensembles de connaissances rend probable l'extraction des autres ensembles de connaissances

parce que tous sont coordonnés à la même représentation mentale, par exemple à une représentation linguistique. Le schéma suivant résume cette idée.

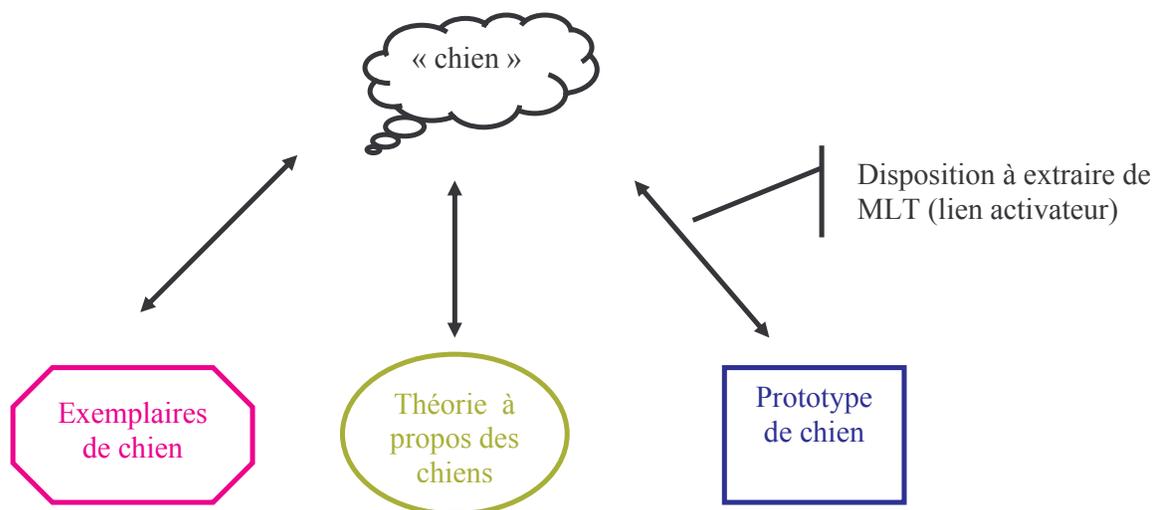


Schéma 8.2.

C'est l'un des nombreux cas où le rejet de l'hypothèse EN au profit de l'HH révèle un nouvel objet de recherche.

### 3.3. L'argument de la sélection des propriétés

#### 3.3.1. L'argument

Un prototype encode des connaissances sur les propriétés qui sont crues être possédées typiquement par les membres de la catégorie dénotée. Un exemplaire encode des connaissances sur les propriétés de certains membres d'une catégorie (chapitre 3). Mais de toute évidence les prototypes n'encodent pas des connaissances sur *toutes* les propriétés typiques des catégories dénotées. Tous les chiens que nous avons rencontrés pèsent moins d'une tonne, moins de deux tonnes et ainsi de suite. Mais un prototype de chien n'encode sans doute aucune connaissance sur ces propriétés. Ce problème n'est pas limité aux propriétés négatives. Tous les chiens vivent sur Terre. Mais un prototype de chien n'encode sans doute aucune connaissance sur cette propriété. Pareillement, les exemplaires n'encodent sans doute pas de connaissances sur *toutes* les propriétés des membres qu'ils dénotent. Médor pèse moins d'une tonne, mais un exemplaire dénotant Médor n'encode sans doute pas le fait que Médor a une telle propriété. Cela soulève la question suivante (voir, par exemple, Armstrong et al. 1983, Murphy et Medin 1985/2004, Medin et Ortony 1989, Gelman 1990, 2001, Murphy

2002) : comment les propriétés représentées par les prototypes et/ou les exemplaires sont-elles déterminées ? Pourquoi la propriété *aboie* est-elle représentée par un prototype de chien ou par les exemplaires qui dénotent la catégorie des chiens, tandis que la propriété *pèse moins d'une tonne* ne l'est sans doute pas ? J'appelle ce problème *le problème de la sélection des propriétés*.

Gelman a particulièrement insisté sur ce point (1990, 2001 – Murphy et Medin 1985/2004 furent, à ma connaissance, les premiers à souligner cette idée). Elle soutient que la TCC répond à cette question : la partie théorique d'un concept détermine à quelles propriétés l'agent prête attention et donc quelles propriétés sont encodées par la partie prototypique et la partie encodant des connaissances à propos d'individus singuliers. Un prototype de chien et un exemplaire de Médor encodent des connaissances sur certaines propriétés plutôt que sur d'autres, parce que nos connaissances théoriques attirent l'attention des agents sur ces propriétés.

Selon la TCC, les différentes parties des concepts sont en effet hiérarchisées. L'agent sait donc que ces différents ensembles de connaissances sont à propos de la même catégorie. Sans quoi, la partie théorique ne pourrait servir de critère pour évaluer l'accomplissement des tâches cognitives au moyen des autres ensembles de connaissances. En conséquence, les connaissances théoriques peuvent déterminer quelles propriétés sont encodées par la partie prototypique et par la partie à propos de certains membres de la catégorie. Au contraire, l'HH n'implique aucune relation particulière entre les différentes représentations mentales, quoique elle ne soit pas incompatible avec l'existence de telles relations (voir ci-dessus). Si c'est le cas, pourquoi nos connaissances théoriques détermineraient-elles les propriétés représentées par les autres concepts ?

### 3.3.2. Réponse

L'argument repose sur un problème réel : puisque les prototypes et les exemplaires ne représentent pas toutes les propriétés de ce qu'ils dénotent, comment les propriétés représentées sont-elles sélectionnées ?

Il est plausible que des connaissances théoriques déterminent les propriétés qui sont représentées lors de la formation du prototype ou de l'ensemble d'exemplaires. Mais cela ne supporte pas la TCC. En effet, il faut distinguer ce que j'ai appelé les théories d'arrière-plan, par exemple la biologie naïve ou l'ontologie naïve, des mini-théories, c'est-à-dire des ensembles de connaissances théoriques sur des catégories particulières (voir le chapitre 3, section 1.4.). Un prototype de chien encode des connaissances sur certaines propriétés

physiques et comportementales des chiens (*mutatis mutandis* pour les exemplaires). Un prototype de table encode des connaissances sur la fonction et peut-être sur la forme des tables (*mutatis mutandis* pour les exemplaires). La différence entre ces deux prototypes peut dériver de nos théories d'arrière-plan sur les animaux et sur les artéfacts. Notre biologie naïve spécifie sans doute que les propriétés physiques et les comportements sont des propriétés importantes des espèces animales : nous prêtons donc attention à ce type de propriétés et nous encodons leur fréquence. La théorie naïve des artéfacts dirige notre attention vers d'autres propriétés. Mais cela ne montre pas que mes connaissances théoriques à *propos des chiens* déterminent quelles propriétés sont encodées par mes connaissances sur les propriétés typiques des chiens et par mes connaissances sur certains chiens en particulier. L'argument proposé ci-dessus repose donc sur une confusion entre *nos théories d'arrière-plan* et *nos mini-théories*.

En outre, dans quelle mesure mes connaissances théoriques déterminent-elles vraiment quelles propriétés sont représentées par les prototypes et les exemplaires ? Par exemple, dans quelle mesure ma théorie d'arrière-plan sur le domaine des artéfacts détermine-t-elle les propriétés de mon prototype de voiture et de l'exemplaire de l'Audi TT que je vois tous les matins devant mon bureau ? Keil a soutenu que les théories d'arrière-plan sont en fait fort vagues (« shallow », voir Keil 1994, Keil et Richardson 1999, voir aussi Atran 1994).<sup>13</sup> L'idée est que le contenu des théories d'arrière-plan est très limité. Supposons ainsi que notre biologie naïve spécifie que les animaux ont des propriétés physiques et des comportements particuliers. Cela ne résout guère le problème de la sélection des propriétés. Les animaux ont de très nombreuses propriétés physiques et comportementales et la plupart ne sont pas représentées par nos prototypes et nos exemplaires d'animaux. Notre biologie naïve ne spécifie sans doute pas que différentes propriétés physiques et comportementales caractérisent les tigres et les mouches. Pourtant, nos prototypes de tigres et de mouches encodent différentes propriétés (*mutatis mutandis* pour les exemplaires).

Enfin, nos prototypes et nos exemplaires peuvent encoder des connaissances sur ce type de propriétés précisément parce que nos connaissances théoriques sont silencieuses à leur propos – aussi bien nos théories d'arrière-plan que nos mini-théories. Par exemple, mon prototype de Porsche encode sans doute le fait que les Porsche ont des roues larges et

---

<sup>13</sup> Récemment, certains psychologues et certains philosophes ont suggéré que les phénomènes que l'on avait expliqués au moyen des théories d'arrière-plan sont peut-être l'effet de biais cognitifs et perceptuels de bas niveau. Par exemple, les phénomènes qui ont suggéré l'existence d'une ontologie naïve (par exemple, Spelke 1990, 1994, Spelke et al. 1992, Carey et Spelke 1994) peuvent être l'effet de biais attentionnels (Leslie et al. 1998, Prinz présentation à la conférence sur l'innéisme mars 2002, Maryland).

l'exemplaire de l'Audi TT devant mon bureau encode le fait que l'aile droite de cette voiture est en mauvais état. Et cela précisément parce que ni la théorie d'arrière-plan des artefacts ni mes mini-théories des Porsche, des Audi TT ou des voitures ne dirigent mon attention vers la taille des pneus de voitures ou vers l'état de leurs ailes droites. C'est la suggestion de Keil (1989, 22-23) :

« Carey and her associates' view of conceptual change seems to be based on the notion that theory completely infiltrates conceptual spaces and that all entities are therefore contrasted on theoretical grounds and no others. This bias may derive from the assumption that concepts are fully specified by their relations to the theories in which they are embedded. (...) But this view seems to rule out any atheoretical storing of information in a domain ; we can only know what our theories interpret for us. It may be, however, that in all domains we also know a great deal more about relations between elements, correlations, and frequencies that have not yet been fully interpreted by the best, most relevant theory. (...)

Consider an example involving cars. A theory of why cars exist, what they are used for, and how subtypes are designed for specific environments tells me why there are typical feature correlations uniquely associated with car subtypes such as sports cars, (...) and off-road vehicles. Thus, I have a coherent cluster of causal beliefs about why the features of low ground clearance, wide wheelbase, powerful high-rpm engine, and two seats frequently cooccur in sports cars and why features such as high ground clearance, short wheelbase, four-wheel drive, and roll bars cooccur with off-road vehicles. Within any of these subclasses I have few if any theoretical or causal beliefs that I use to interpret correlations ; yet, I daily use those correlations to distinguish, for example, a Ford from a Chevy sedan. Theory may tell me to look at correlations between trim and hood shape and ignore license plate, the color, and the presence of fuzzy dice ; but beyond telling me to look for correlations in some domains and not others, it does not tell me which ones to look for and what their values are likely to be. This atheoretical knowledge is part of the concept because it provides a framework for further theory development ; it is the structure on which future causal beliefs will be overlaid. »

Je conclus que l'argument de la sélection des propriétés n'est pas convaincant. En résumé, il est probable que nos théories d'arrière-plan dirigent notre attention vers différents

types de propriétés dans différents domaines. Elles déterminent donc partiellement certaines propriétés encodées par les prototypes et les exemplaires. Cependant, elles ne spécifient pas en détail quelles propriétés doivent être encodées. Au contraire, il se peut que nos prototypes et exemplaires encodent précisément certaines propriétés qui ne sont pas spécifiées par nos connaissances théoriques. Enfin, dans la mesure où les théories d'arrière-plan et les mini-théories sont différentes, la sélection des propriétés représentées par les prototypes et les exemplaires ne vient pas à l'appui de la TCC.

### *3.4. Conclusion*

Résumons la discussion menée dans cette section. J'ai cherché à montrer que différentes considérations qui viennent de prime abord à l'appui de la TCC ne viennent en fait pas à son appui. Nous distinguons bien les apparences de la réalité, nos ensembles de connaissances sont bien coordonnés et nos prototypes et exemplaires représentent bien des propriétés spécifiques. Je n'ai contesté aucun de ces trois phénomènes. Mais j'ai montré qu'ils ne venaient pas à l'appui de la TCC. Nous distinguons la réalité de l'apparence quand nous employons nos concepts essentialistes et théoriques, mais certaines intuitions sémantiques montrent que nous sommes aussi prêts à tenir la ressemblance avec un prototype ou certains exemplaires pour le critère de correction de nos jugements. La coordination de nos ensembles de connaissances à propos d'une même catégorie n'est qu'un cas particulier de la coordination de nos concepts. Et la sélection des propriétés n'est pas expliquée par la TCC. Je conclus donc que ces phénomènes ne nous autorisent pas à préférer la TCC à l'hypothèse de l'hétérogénéité.

Bien plus, l'existence d'intuitions sémantiques apparemment contradictoires et les phénomènes décrits dans le chapitre 7 favorisent l'HH. Ils montrent que les ensembles de connaissances à propos d'une même catégorie sont bien fonctionnellement indépendants. Nous sommes prêts à approuver à la fois (1) et (2) (voir ci-dessus) parce que le prototype de grand-mère (ou un ensemble d'exemplaires) et le concept essentialiste de grand-mère sont fonctionnellement indépendants. Je conclus donc que si les théories composites des concepts sont motivées par l'idée que les différents ensembles de connaissances à propos d'une même catégorie sont hiérarchisés, l'HH leur est préférable.

## **4. La seconde branche du dilemme**

Dans la dernière section de ce chapitre, j'examine la seconde branche du dilemme présenté ci-dessus. La TCC soutient qu'un concept encode trois types de connaissances, des connaissances à propos d'instances particulières d'un concept, des connaissances à propos des propriétés typiques de la catégorie et des connaissances théoriques. J'ai montré dans la section précédente que nous n'avons aucune raison de croire que ces trois types de connaissances sont hiérarchisés. Supposons maintenant que les avocats de la TCC admettent ce point, mais qu'ils veuillent quand même maintenir que ces trois ensembles de connaissances ne forment pas trois concepts différents, mais trois parties d'un même concept. Si les avocats de la TCC adoptent une telle position – c'est-à-dire embrassent la seconde alternative du dilemme présenté dans la première section, je réponds que la différence entre la TCC et l'HH est alors purement terminologique.

#### *4.1. Deux variantes terminologiques*

Les avocats des théories composites des concepts ne peuvent soutenir que les concepts encodent un type spécifique de connaissances ou supportent des processus cognitifs spécifiques – c'est-à-dire, ont des propriétés sémantiques et fonctionnelles spécifiques (voir le chapitre 4). Ils doivent soutenir que chaque partie d'un concept encode un type particulier de connaissances et possède des propriétés fonctionnelles spécifiques.

En outre, par hypothèse, les avocats de la TCC ont reconnu que les ensembles de connaissances à propos d'une même catégorie sont fonctionnellement indépendants. Les avocats de la TCC doivent donc soutenir que les relations entre les parties des concepts sont identiques aux relations entre différentes représentations mentales. En d'autres termes, le fait que certaines connaissances sur les propriétés typiques des chiens, certaines connaissances sur des chiens particuliers et certaines connaissances théoriques à propos des chiens forment des parties d'un même concept ne se manifeste par aucune relation particulière entre ces ensembles de connaissances. Les relations qui peuvent exister entre des représentations mentales coréférentielles peuvent exister entre les différentes parties d'un même concept – et inversement.

En conséquence, un avocat de l'HH peut traduire dans son propre vocabulaire les thèses qu'un avocat de la TCC peut avancer. Considérons quelques exemples.

1. Les concepts sont en général divisés en trois parties.
- 1'. Les catégories sont en général représentées par trois concepts coréférentiels.

2. Une partie encode des connaissances sur des individus particuliers, une autre, des connaissances sur les propriétés typiques de la catégorie et la dernière, des connaissances théoriques.
- 2'. Un type de concepts encode des connaissances sur des individus particuliers, un autre, des connaissances sur les propriétés typiques de la catégorie et le dernier, des connaissances théoriques.
3. Une partie des concepts est utilisée dans des processus qui reposent sur la ressemblance additive, une autre, dans des processus qui reposent sur la ressemblance multiplicative et une dernière, dans des processus qui reposent sur des règles (voir le chapitre 4, section 2.).
- 3'. Un type de concepts est utilisé dans des processus qui reposent sur la ressemblance additive, un autre, dans des processus qui reposent sur la ressemblance multiplicative et le dernier, dans des processus qui reposent sur des règles.

Cela montre que la différence entre les deux théories n'est pas substantielle : elles décrivent les mêmes faits. La différence est donc purement *terminologique*.

#### 4.2. Le « bon niveau » de la recherche psychologique

Il faut aussi remarquer que selon cette version de la TCC, la recherche psychologique ne porte pas sur les concepts eux-mêmes, mais sur leurs parties. Car les généralisations sur les ensembles de connaissances utilisés par défaut dans nos processus cognitifs portent alors sur les parties des concepts. Un avocat de cette version de la TCC pourrait, par exemple, poser les questions suivantes : quelles sont les propriétés de la partie des concepts qui encode des connaissances à propos de membres particuliers d'une catégorie ? Quelles sont les propriétés des processus cognitifs qui font appel à cette partie ? etc. Puisque les parties d'un même concept n'ont pas, par hypothèse, de relations spécifiques entre elles, le fait que les ensembles de connaissances soient des parties d'un même concept n'introduit aucune généralisation supplémentaire qui porterait sur les concepts eux-mêmes et non sur leurs parties.

Or, c'est là le point crucial du rejet de l'hypothèse Espèce Naturelle (chapitre 2). Rejeter l'hypothèse EN, c'est fondamentalement affirmer que « concept » n'est pas une notion utile pour formuler des généralisations empiriques en psychologie. Si les avocats de la TCC rejettent la hiérarchisation fonctionnelle des parties des concepts, ils admettent aussi cette idée. En conséquence, tout comme les avocats de l'HH, ils rejettent l'hypothèse EN. Ils soutiennent de fait que le travail empirique en psychologie des compétences cognitives

supérieures ne doit pas être fait au niveau des concepts. C'est là la thèse principale que je soutiens dans ce travail.

#### *4.3. Quelle variante terminologique préférer ?*

Si cette version de la TCC et l'HH ne sont que deux variantes terminologiques, on peut à juste titre poser la question suivante : pourquoi préférer l'HH à cette version de la TCC ? Pourquoi dire que les concepts se divisent en plusieurs types plutôt que les concepts ont différentes parties ?

Tout d'abord, s'il était clair que cette version de la TCC n'impliquait pas la hiérarchisation des différents ensembles de connaissances à propos d'une même catégorie, il n'y aurait aucune raison de préférer une variante à l'autre. Mais précisément, parce que cela n'est pas clair, il est préférable de soutenir l'HH à cette variante. En effet, comme nous l'avons vu plus haut (voir la section 2), la TCC est associée à la hiérarchisation des différents ensembles de connaissances à propos d'une même catégorie. Il vaut donc mieux éviter toute ambiguïté et soutenir l'HH plutôt que cette version de la TCC.

#### *4.4. Conclusion*

Dans cette dernière section, j'ai défendu la seconde branche du dilemme posé dans la première section de ce chapitre. Si les avocats de la TCC renoncent à la hiérarchisation fonctionnelle des différentes parties des concepts, leur théorie et l'HH ne diffèrent que terminologiquement. Les avocats de l'HH peuvent traduire dans leurs propres termes les généralisations que les avocats de cette version de la TCC pourraient proposer. Et tous s'accordent pour rejeter en fait l'hypothèse EN : la notion de concept est inappropriée pour formuler des généralisations à propos de nos compétences cognitives supérieures.

### **5. Conclusion**

L'objet de ce travail de thèse est de critiquer un postulat admis par de nombreux psychologues des concepts et des compétences cognitives supérieures : les concepts forment une espèce naturelle, autrement dit, il est possible de formuler de nombreuses généralisations empiriques à leur propos. Je soutiens que les psychologues doivent renoncer à ce postulat et formuler explicitement leurs généralisations à propos de types particuliers de concepts, à savoir (entre autres) les théories, les exemplaires et les prototypes. Pour défendre cette position, j'ai montré

que de nombreuses données empiriques supportaient l'hypothèse de l'hétérogénéité, à savoir la thèse que les concepts se divisent en types qui ont peu en commun.

Ce faisant, j'ai noté qu'une théorie composite particulière, postulant que les concepts sont constitués d'une partie encodant des connaissances sur les propriétés typiques des catégories dénotées, d'une partie encodant des connaissances théoriques à leur propos et d'une partie encodant des connaissances sur les propriétés de membres particuliers de ces catégories, expliquait aussi bien les processus cognitifs que les approches fondées sur l'HH.

Ce chapitre a donc pour objet de critiquer cette TCC. J'ai soutenu que les avocats de la TCC devaient expliquer en quoi les différents ensembles de connaissances à propos d'une même catégorie formaient des parties d'un même concept. J'ai proposé que cela signifiait que ces parties étaient fonctionnellement hiérarchisées : l'une d'entre elles est supposée fonctionner comme le critère de correction de l'accomplissement des tâches cognitives. J'ai cependant montré que peu de données empiriques venaient à l'appui de cette idée. Au contraire, certaines intuitions linguistiques suggèrent que les ensembles de connaissances sont fonctionnellement indépendants. Finalement, j'ai montré que si les avocats de la TCC rejettent la hiérarchisation fonctionnelle des parties des concepts, leur position et l'HH sont de simples variantes terminologiques.

Je conclus donc que les explications des processus cognitifs qui reposent sur l'HH fournissent la meilleure explication actuelle des propriétés des processus cognitifs. Ce chapitre termine donc la défense de la prémisse 2 de l'argument de la nécessité explicative (ANE) : il est justifié de postuler plusieurs types de représentations mentales (à savoir, entre autres, les prototypes, les exemplaires et les théories) afin d'expliquer le déroulement des processus cognitifs supérieurs. Je conclus donc qu'il faut rejeter l'hypothèse EN : les concepts ne forment pas une espèce naturelle.

## Conclusion

---

L'essentiel de ce travail consiste en une discussion détaillée, et, je l'espère, convaincante, des prémisses de l'argument de la nécessité explicative, particulièrement de la seconde prémisse : il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs. Pour ce faire, j'ai longuement présenté et discuté de nombreux résultats expérimentaux de la psychologie expérimentale des concepts et des fonctions cognitives supérieures. Cette approche n'est pas sans risque : à regarder de trop près les détails expérimentaux, on risque de perdre de vue l'idée principale. Pour que les arbres ne cachent pas la forêt, il est donc temps de prendre un peu de hauteur. Par ailleurs, il importe maintenant d'examiner les conséquences du rejet de l'hypothèse EN, à savoir de la thèse selon laquelle les concepts forment une espèce naturelle. Qu'implique ce rejet pour les psychologues ? Comment devraient-ils étudier les représentations mentales qui sont utilisées par défaut dans nos processus cognitifs supérieurs ? Cette conclusion s'attelle à ces tâches.

Dans la première section, je résume la thèse centrale de ce travail et les principaux moments de l'argumentation. Dans la seconde section, je présente les conséquences négatives et critiques du rejet de l'hypothèse EN. En d'autres termes, je réponds aux questions suivantes : quels aspects de l'activité des psychologues des concepts et des fonctions cognitives supérieures sont critiquables, étant donné le caractère erroné de l'hypothèse EN ? Et en quoi ? Dans la troisième section, je discute les conséquences positives du rejet de l'hypothèse EN. Je cherche à répondre à la question suivante : quels sont les apports positifs de ce travail ? Je montre que le rejet de l'hypothèse EN rend saillantes des questions qui n'ont pas reçu l'attention qu'elles méritent. Finalement, dans la dernière section, je cherche à répondre aux questions suivantes : si la notion de concept ne désigne pas une espèce naturelle, doit-on la conserver ? A-t-elle un rôle à jouer dans une psychologie scientifique ? Faut-il au contraire l'éliminer ? Si oui, cela implique-t-il que nous n'avons pas de concept ?

### **1. Les concepts ne forment pas une espèce naturelle**

#### *1.1. L'hypothèse EN*

Dans la psychologie contemporaine des concepts et des fonctions cognitives supérieures, particulièrement dans la psychologie expérimentale qui s'est développée depuis les années 70,

le terme « concept » est un terme théorique, introduit pour désigner des entités théoriques particulières. Il désigne des ensembles de connaissances à propos de classes d'entités qui sont stockés dans la mémoire à long terme et qui sont utilisés par défaut dans nos processus cognitifs supérieurs, par exemple quand nous catégorisons, induisons, faisons une analogie, comprenons des énoncés linguistiques etc. L'objet de la psychologie des concepts est de décrire les propriétés qui sont communes à ces ensembles de connaissances ou, au moins, à la plupart d'entre eux. Les psychologues étudient ainsi les questions suivantes :

- (i) Quel type d'informations est encodé par les concepts ? Ces informations sont-elles générales ? Sont-elles perceptuelles ? etc.
- (ii) Comment les concepts sont-ils utilisés dans les processus cognitifs supérieurs ?
- (iii) Comment les concepts sont-ils réalisés dans le cerveau ? etc.

Les psychologues des concepts admettent donc que, par-delà les différences entre concepts (un concept de chien est différent d'un concept de chat) et entre types de concepts (les concepts d'animaux ont des propriétés spécifiques qui les distinguent des concepts d'artéfacts), il existe un certain nombre de propriétés communes à l'ensemble des concepts. En conséquence, avec la classe des concepts, ils croient avoir identifié l'une de ces classes d'entités qui sont essentielles pour les sciences empiriques, à savoir une espèce naturelle. Selon les psychologues des concepts et selon de nombreux philosophes de la psychologie, les concepts forment une classe d'entités qui donnent lieu à de nombreuses projections inductives : étudier les propriétés de certains concepts permet de mettre au jour les propriétés de tous les concepts. Classifier nos états mentaux au moyen de la notion de concept est donc important pour la psychologie scientifique. Le terme « concept » joue ainsi un rôle essentiel en psychologie.

### *1.2. Le rejet de l'hypothèse EN*

L'objet de ce travail est de critiquer la thèse selon laquelle les concepts forment une espèce naturelle. Je soutiens en effet que l'image de notre cognition supérieure qui est admise par la plupart des psychologues est erronée. La classe des concepts, c'est-à-dire la classe des représentations mentales qui sont utilisées par défaut dans nos processus cognitifs supérieurs, n'est pas une espèce naturelle : ce n'est pas l'une de ces classes qui permettent de faire de nombreuses inductions. La classe des concepts ne peut remplir les rôles qui lui reviennent

selon les psychologues des concepts et des fonctions cognitives supérieures et selon la plupart des philosophes de la psychologie.

En effet, je soutiens que la classe des concepts se divise en fait en sous-classes qui ont fort peu en commun. Une même classe d'objets, par exemple, les chiens, est représentée par plusieurs concepts qui appartiennent à ces classes fort différentes de représentations mentales. J'ai appelé cette thèse *l'hypothèse de l'hétérogénéité*. Il est plausible que ces classes soient elles-mêmes des espèces naturelles. En lieu et en place de la classe des concepts, ce sont donc les classes pertinentes pour étudier de manière empirique les représentations mentales qui figurent par défaut dans nos processus cognitifs supérieurs. Une classification de nos représentations mentales au moyen de ces classes doit remplacer, pour l'étude scientifique de la cognition, la classification actuelle, qui distingue seulement les concepts des autres types de représentations mentales. En conséquence, la notion de concept est inappropriée si l'on cherche à formuler des généralisations à propos des ensembles de connaissances qui sont utilisés par défaut dans nos processus cognitifs supérieurs. Il faut la remplacer par des notions dénotant les sous-classes de concepts mentionnées ci-dessus.<sup>1</sup>

J'ai en outre proposé que les principaux paradigmes en psychologie des concepts, à savoir le paradigme postulant des prototypes, le paradigme postulant des exemplaires et le paradigme postulant des théories avaient identifié certains de ces types de concepts. En d'autres termes, je soutiens que la classe des concepts se divise, entre autres, en prototypes, exemplaires et théories. De nombreuses classes d'objets sont telles que nous avons en mémoire à long terme un prototype, un ensemble d'exemplaires et une théorie qui encodent des connaissances à leur propos.<sup>2</sup> Un prototype est un ensemble de connaissances statistiques sur les propriétés que possèdent les membres de la classe dénotée. Un ensemble d'exemplaires est un ensemble de représentations de membres de la classe dénotée. Une théorie est un ensemble de connaissances qui ressemble à une théorie scientifique, dans la mesure où il contient des connaissances nomologiques, causales ou fonctionnelles à propos de la classe dénotée. Le prototype, l'ensemble d'exemplaires et la théorie à propos d'une classe d'objets ne sont pas des parties d'un unique concept, mais bien des représentations mentales différentes.

### *1.3. L'argument de la nécessité explicative*

---

<sup>1</sup> J'examine ci-dessous la question de l'élimination intégrale de cette notion.

<sup>2</sup> Notons ici que je ne soutiens pas que *toutes* les classes d'entités soient représentées par un exemplaire, un prototype et une théorie.

Pourquoi rejeter l'hypothèse EN au profit de l'hypothèse de l'hétérogénéité ? L'argument de la nécessité explicative soutient qu'il est nécessaire de postuler plusieurs types de représentations mentales, entre autres, des prototypes, des exemplaires et des théories, afin d'expliquer le déroulement des processus cognitifs. Comme les concepts sont précisément ces représentations mentales qui expliquent le déroulement des processus cognitifs, la classe des concepts se divise, entre autres, en prototypes, exemplaires et théories. Comme ces trois types de représentations mentales ont peu en commun, la classe des concepts ne forme pas une espèce naturelle, mais se divise en types de représentations mentales qui ont peu en commun.

La prémisse 1 de l'argument soutient que les concepts sont les représentations mentales dont les propriétés expliquent le déroulement des processus cognitifs supérieurs. J'ai cherché à défendre cette prémisse dans le premier chapitre de ce travail, consacré à la présentation de la notion psychologique de concept et de la psychologie des concepts en général. Bien que cette thèse soit admise par la quasi-totalité des psychologues, elle est cependant fort controversée. En effet, Fodor a rejeté de manière répétée l'idée que les propriétés des concepts expliquent les propriétés des processus cognitifs. En bref, j'ai proposé que Fodor ne pouvait rejeter cette prémisse qu'en sélectionnant de manière arbitraire certains phénomènes empiriques parmi l'ensemble des phénomènes à expliquer.

La prémisse 2 de l'argument de la nécessité explicative soutient qu'il est justifié de postuler plusieurs types de représentations mentales afin d'expliquer le déroulement des processus cognitifs supérieurs. C'est la prémisse centrale de cet argument. Il est impossible d'argumenter *a priori* pour cette prémisse. Pour ce faire, il est au contraire nécessaire de prendre en compte les propriétés des processus cognitifs qui sont découvertes expérimentalement. On peut alors montrer que la meilleure explication de ces propriétés postule plusieurs types de représentations, à savoir des prototypes, des exemplaires et des théories. Les chapitres 5, 6, 7 et 8 ont pour objet de défendre la seconde prémisse. Dans le chapitre 5, j'ai cherché à montrer que l'explication des propriétés de la catégorisation venait à son appui. Plutôt que de chercher un unique type de concepts, figurant dans un unique processus de catégorisation, il semble nécessaire de postuler plusieurs types de concepts qui sont utilisés dans des processus cognitifs différents. J'ai appelé l'idée selon laquelle une compétence cognitive comme la catégorisation repose sur plusieurs processus cognitifs la théorie pluraliste de la cognition (voir la section 3.). Dans le chapitre 6, j'ai examiné la littérature empirique sur la composition des concepts, c'est-à-dire sur notre capacité de produire des concepts nouveaux à partir des concepts stockés en mémoire à long terme. J'ai proposé que cette capacité repose sur un mécanisme cognitif dont les inputs incluent des

prototypes, des exemplaires et des théories. Dans le chapitre 7, j'ai examiné l'usage du terme « eau ». J'ai proposé que les usages apparemment incompatibles du terme « eau » peuvent être expliqués si l'on supposait que ce terme est associé à plusieurs concepts. J'ai proposé que nous avons au moins un concept théorique et essentialiste d'eau et un concept non théorique d'eau, probablement un prototype ou un ensemble d'exemplaires. Postuler plusieurs types de concepts, à savoir des prototypes, des exemplaires et des théories, permet donc d'expliquer de nombreux phénomènes à propos des processus cognitifs supérieurs. Au contraire, on ne peut expliquer l'ensemble de ces phénomènes en postulant seulement des prototypes, des exemplaires ou des théories. On pourrait toutefois objecter qu'une théorie composite des concepts est à même de rendre compte de ces phénomènes. Supposons en effet que les concepts soient composés de plusieurs parties : l'une d'entre elles pourrait encoder les connaissances que les prototypes sont supposés encoder, une autre les connaissances que les ensembles d'exemplaires sont supposés encoder et une troisième les connaissances que les théories sont supposées encoder. Dans le chapitre 8, je critique ce type de théories composites. Les connaissances théoriques, les connaissances statistiques et les connaissances sur les individus ne sont pas essentiellement hiérarchisées, ce qui s'accorde mal avec ces théories composites.

La prémisse 3 de l'argument de la nécessité explicative soutient que les prototypes, les exemplaires et les théories ont peu de propriétés en commun. Le chapitre 3 et 4 ont pour objet de défendre cette prémisse. Dans le chapitre 3, je présente ces trois entités théoriques. Dans le chapitre 4, je montre qu'elles sont supposées avoir des propriétés sémantiques et fonctionnelles différentes. Etant donné les propriétés des représentations mentales qui importent pour les psychologues et les philosophes de la psychologie, je conclus que les prototypes, les exemplaires et les théories forment trois types de représentations mentales qui ont peu en commun.

Je conclus donc que peu de propriétés sont communes aux représentations qui expliquent le cours des processus cognitifs supérieurs. Dans le chapitre 2, j'ai expliqué en détail la notion d'espèce naturelle. Je favorise la notion causale d'espèce naturelle formulée par Boyd et développée par Griffith. Elle s'applique à de très nombreuses classes sans être vide et elle peut aussi s'appliquer aux classes d'entités psychologiques. Je conclus que les concepts ne forment pas une espèce naturelle.

#### *1.4. Remarques*

L'argument de la nécessité explicative est de toute évidence *empirique*. La prémisse centrale de cet argument, à savoir la prémisse 2, ne peut être établie qu'au moyen de données empiriques à propos de nos processus cognitifs. De ce fait, la solidité de la conclusion dépend de la validité des données empiriques sur lesquelles repose cet argument. Tout comme les théories qu'elles confirment ou invalident, les données empiriques sont susceptibles d'être rejetées. Elles ne forment pas des fondations infalsifiables de notre connaissance empirique. Il est donc possible que certaines données empiriques résumées dans ce travail soient rejetées par la psychologie des décades à venir ou que leur statut et leur pertinence soient réinterprétés. De ce fait, le rejet de l'hypothèse EN pourrait lui-même devoir être révisé, si le progrès de la psychologie conduisait les psychologues à mettre en doute les données empiriques discutées dans ce travail. Quoique cela soit possible, je soutiens que cela est peu probable. Ces phénomènes sont extrêmement robustes et ont été répliqués à de nombreuses reprises.

En outre, on pourrait espérer formuler une théorie des concepts qui soit à même de rendre compte de l'ensemble des données empiriques présentées dans ce travail en ne postulant qu'un type de concepts. Certes, les théories qui ne postulent que des prototypes, des exemplaires ou des théories ne sont pas à même de rendre compte des phénomènes présentés ici. Certes, les théories composites ne sont pas satisfaisantes. Mais je n'ai pas montré qu'il était vain d'espérer rendre compte de ces phénomènes au moyen d'un type unique de concepts. Il se peut que nous manquions seulement d'imagination pour formuler la bonne théorie des concepts. En réévaluant les données empiriques et en proposant une nouvelle théorie des concepts, les psychologues du vingt-et-unième siècle établiront peut-être la validité de l'hypothèse EN.

De nouveau, cela n'est pas impossible. La situation dialectique est en effet la suivante. Nous connaissons un grand nombre de phénomènes empiriques à propos des concepts et des processus cognitifs conceptuels. Les théories actuelles des concepts ne sont pas à même de rendre compte de tous ces résultats empiriques sur les concepts. Deux options sont possibles. On peut conclure qu'une nouvelle théorie des concepts, postulant un type homogène de représentations, est requise pour rendre compte de l'ensemble de ces données expérimentales (ou de la plupart d'entre elles). De nombreux psychologues des concepts et des fonctions cognitives supérieures embrassent cette option. On peut conclure au contraire que l'hypothèse EN est fautive : il est vain de chercher à rendre compte de tous ces phénomènes en posant une classe homogène de représentations mentales. Au contraire, on peut expliquer ces phénomènes si l'on admet l'existence de types de concepts qui ont peu à voir les uns avec les

autres, c'est-à-dire si l'on admet l'hypothèse de l'hétérogénéité. Différents types de concepts rendent compte de différents phénomènes. Certains psychologues ont choisi cette option (Smith et al. 1998, Smith et Jonides 1999 etc.). Ils explorent des problèmes que l'acceptation inconditionnelle de l'hypothèse EN avait conduit à négliger (voir la section 3. de ce chapitre).

Toutefois, ces deux options ne sont pas équivalentes. Même si l'on ne peut exclure la formulation d'une nouvelle théorie des concepts, qui postulerait une classe homogène de représentations et qui soit à même de rendre compte de la plupart des phénomènes empiriques, l'hypothèse de l'hétérogénéité est plus plausible. En effet, si l'on accepte cette hypothèse, on est à même d'expliquer l'ensemble des phénomènes qui sont si problématiques pour les théories contemporaines des concepts. On n'a pas à parier sur l'invention d'une théorie qui serait à même d'expliquer ces phénomènes. Entre une explication qui explique *présentement* les *explananda* et une explication possible qui pourrait *éventuellement* expliquer ces mêmes *explananda*, il est raisonnable de choisir la première option.

### *1.5. Conclusion*

Dans cette section, j'ai résumé la thèse principale de ce travail, le rejet de l'hypothèse EN au nom de l'hypothèse de l'hétérogénéité. J'ai brièvement résumé l'argument de la nécessité explicative et j'ai insisté sur le caractère empirique et falsifiable de cet argument. Il faut conclure que la classe des concepts ne joue pas le rôle crucial dans les théories de la cognition supérieure qu'elle est censée jouer selon de nombreux psychologues. Selon eux, c'est une classe qui donne lieu à de nombreuses généralisations scientifiques. Ils supposent aussi que les propriétés générales des concepts expliquent de nombreuses propriétés des processus cognitifs supérieurs (voir le chapitre 1, section 2.3.2.). Si les concepts ne forment pas une classe naturelle, la notion de concept ne joue aucun de ces rôles. Ils sont au contraire remplis par les classes des prototypes, des exemplaires et des théories. Ainsi, le rejet de l'hypothèse EN implique que la notion de concept ne devrait au mieux jouer qu'un rôle marginal comparé aux notions de prototype, d'exemplaire et de théorie.

## **2. Mettre fin à la controverse sur la nature des concepts**

Dans cette section, je tire les conclusions critiques de la position défendue dans ce travail.

### *2.1. La controverse sur la nature des concepts*

A plusieurs reprises, j'ai insisté sur le fait que les principales théories des concepts sont habituellement traitées comme des théories rivales (Smith et Medin 1981, Medin 1989, Komatsu 1992, Hampton 1997a, Laurence et Margolis 1999, Murphy 2002, Prinz 2002, chap. 2-4). Ces théories sont supposées décrire les propriétés principales des concepts. Comme elles postulent des entités qui ont peu à voir les unes avec les autres (chapitre 4), elles sont tenues pour incompatibles.

Dès le début des années 80, cette approche a engendré une controverse entre les avocats des différents paradigmes des concepts. Les psychologues ont cherché à montrer la supériorité de leur paradigme des concepts par rapport aux autres approches. Pour ce faire, ils ont cherché à mettre au jour des phénomènes aisément explicables par leur propre théorie, mais problématiques pour les autres théories (voir le chapitre 5, voir aussi Barsalou 1990 pour des commentaires lucides sur les stratégies de recherche des psychologues des concepts). J'appelle cette controverse *la controverse sur la nature des concepts*.

*Il est temps de mettre fin à cette controverse.* Elle suppose que les principaux paradigmes des concepts ont pour objet de décrire une classe homogène de représentations mentales, en bref une espèce naturelle. Supposons au contraire qu'il existe plusieurs types de concepts qui ont peu à voir les uns avec les autres, comme j'ai cherché à le montrer tout au long de ce travail. Si cela est le cas, il est nécessaire d'embrasser *plusieurs* théories pour décrire ces types de concepts. Ces théories ne sont ni incompatibles ni rivales, dans la mesure où elles ne portent pas sur les mêmes entités.

En outre, j'ai cherché à montrer que le paradigme postulant des prototypes, le paradigme postulant des exemplaires et le paradigme postulant des théories décrivent les principaux aspects de certains de ces types de concepts. En conséquence, les psychologues et les philosophes de la psychologie devraient cesser de traiter ces trois paradigmes comme rivaux. Les psychologues des concepts ont dépensé beaucoup d'énergie et d'intelligence dans la controverse sur la nature des concepts. Comme nous allons le voir dans la section suivante, cette énergie pourrait être réorientée vers de nouveaux problèmes – qui ne deviennent pressants qu'une fois l'hypothèse EN rejetée.

Soulignons que, de toute évidence, l'hypothèse EN n'implique nullement que n'importe quelle théorie est correcte. Tout d'abord, même si les paradigmes postulant des prototypes, des exemplaires et des théories décrivent bien les principales propriétés de trois types de concepts, il existe de nombreuses théories différentes postulant des prototypes, de nombreuses théories postulant des exemplaires et de nombreuses théories postulant des

concepts théoriques (voir la section 3.).<sup>3</sup> Si les prototypes, forment une espèce naturelle, toutes les théories des prototypes ne peuvent être vraies. C'est aussi le cas des exemplaires et des concepts théoriques. Par ailleurs, il est possible qu'il existe d'autres types de concepts – outre les prototypes, les exemplaires et les théories (voir la section 3). La nature de ces types de concepts reste à décrire. La validité des théories que l'on pourrait proposer doit être évaluée empiriquement.

## 2.2. De l'utilité des fausses théories

On peut se demander si l'hypothèse EN et la controverse sur la nature des concepts ont vraiment nui au développement de la psychologie des concepts et de la psychologie des fonctions cognitives supérieures. On ne peut répondre simplement à cette question.

De toute évidence, parce qu'ils ont embrassé l'hypothèse EN, nombre de psychologues ont négligé l'approche des concepts que je tiens pour de loin la plus plausible, à savoir l'hypothèse de l'hétérogénéité. Ils n'ont donc pu développer une classification adéquate des représentations mentales qui sont utilisées par défaut dans nos processus cognitifs supérieurs. C'est là de toute évidence une conséquence regrettable de l'hypothèse EN.

Par ailleurs, la controverse sur la nature des concepts a conduit les psychologues à multiplier les hypothèses *ad hoc*. En effet, comme les psychologues ont supposé pendant trente ans que les concepts formaient une classe homogène d'états mentaux, ils ont admis le principe selon lequel les entités qu'ils postulaient devaient être à même de rendre compte de tous les phénomènes pertinents – ou au moins de la plupart d'entre eux. La seule manière de pouvoir espérer rendre compte de la diversité des phénomènes cognitifs concernant les concepts consiste à recourir à des hypothèses *ad hoc* (Barsalou 1990). Si les hypothèses *ad hoc* ne sont pas par principe condamnables (Kuhn 1970), le recours systématique à ces hypothèses est néanmoins problématique. Les théories des exemplaires développées par Nosofsky illustrent ce problème (Nosofsky 1986, 1988, 1992, 2000, Ashby et Maddox 1993, voir le résumé dans Murphy 2002, 65-71). Le modèle contextuel généralisé (« Generalized Context Model ») proposé dans les années 80 par Nosofsky – qui développe le modèle contextuel (« context model ») de Medin et Schaffer (1978) – est déjà très souple. La distance psychologique entre un objet et un exemplaire<sup>4</sup> dépend d'un paramètre libre,  $w$ , qui correspond à l'attention portée à certaines dimensions. La valeur de ce paramètre est fonction

---

<sup>3</sup> Voir le chapitre 3, section 1. à propos de la différence entre les paradigmes et les théories.

<sup>4</sup> Il faut se représenter un espace mental dont les dimensions sont les propriétés possibles des objets. En fonction de leurs propriétés, les objets occupent des positions particulières dans cet espace et sont plus ou moins éloignés les uns des autres.

du contexte. En outre, cette distance dépend aussi de la métrique utilisée. Le modèle utilise souvent une métrique euclidienne, mais il fait parfois appel, dans certains contextes, à d'autres métriques (en particulier, à la métrique appelée « city block »). Pour transformer la distance psychologique en ressemblance, il faut postuler une autre variable libre, *c*. Cette variable exprime le fait que dans certains contextes, mais non dans tous, la ressemblance est une fonction exponentielle décroissante de la distance. Ces paramètres libres, dont la valeur dépend du contexte, sont autant d'hypothèses *ad hoc* pour rendre compte des données. Dans les années 90, pour rendre compte de l'ensemble des données, en particulier des données naturellement expliquées par les théories des prototypes (Smith et Minda 2000), les partisans du modèle contextuel généralisé ont introduit un autre paramètre libre, *gamma*. Gamma correspond au fait que dans certains cas, les décisions des sujets peuvent être caractérisées de manière déterministe, tandis que dans d'autres cas, elles ne peuvent être caractérisées que de manière probabiliste. Au moyen de ce nouveau paramètre, le modèle semble en mesure de rendre compte d'un très grand nombre de résultats expérimentaux sur la catégorisation d'items dans des catégories d'objets artificiels. Mais, c'est au prix de nombreux paramètres libres. Prétendre expliquer toutes les données expérimentales sur la catégorisation au moyen d'un seul type d'entités semble donc conduire à affaiblir les théories des concepts.

L'hypothèse EN et la controverse sur la nature des concepts ont donc bien eu des conséquences néfastes sur le développement de la psychologie des concepts et des fonctions cognitives supérieures. Cependant, ce jugement doit être nuancé. En effet, par certains aspects, la controverse sur la nature des concepts a mené à de nouvelles découvertes à propos des compétences conceptuelles, comme la catégorisation, le raisonnement inductif ou le raisonnement déductif. En effet, cette controverse a conduit les psychologues à chercher de nouveaux phénomènes empiriques, dans l'espoir d'établir la supériorité de leur paradigme préféré sur les autres paradigmes. Par exemple, les théoriciens des prototypes ont découvert des phénomènes de typicalité au-delà du domaine de la catégorisation. Les travaux de Osherson et collègues sur l'induction et les travaux de Sloman sur la déduction illustrent cette approche (Osherson et al. 1990, Sloman 1998). Les partisans des concepts théoriques ont cherché à trouver des phénomènes inexplicables par les théories postulant des prototypes ou des exemplaires. Les travaux de Susan Gelman sur les inductions des jeunes enfants et les travaux de Keil sur la catégorisation des jeunes enfants en sont les résultats.

De toute évidence, ce n'est pas la première fois que de fausses théories conduisent à des découvertes empiriques (Laudan 1981, mais voir la discussion de Kitcher 1993). Toutefois, le cas de la controverse sur la nature des concepts est particulier. En effet, les trois

principaux paradigmes des concepts décrivent correctement les principales propriétés de trois types de concepts. Cela explique pourquoi la controverse sur la nature des concepts a conduit à tant de découvertes expérimentales sur nos processus cognitifs supérieurs. Les avocats des prototypes ont raison de croire en l'existence de prototypes et de supposer que les prototypes sont utilisés quand nous catégorisons, induisons, déduisons etc. Cela explique pourquoi ils ont pu trouver tant de phénomènes empiriques qui viennent à l'appui de l'existence de prototypes et de leur utilisation dans nos processus cognitifs supérieurs. Ils ont seulement tort de croire que *tous* les concepts sont des prototypes et de chercher à expliquer l'ensemble des phénomènes qui caractérisent nos compétences cognitives supérieures au moyen des prototypes. C'est aussi le cas des exemplaires et des concepts théoriques.

### 2.3. *Les principes d'unification et de parcimonie*

Certains lecteurs ont probablement à l'esprit l'objection suivante : si tant de données empiriques viennent à l'appui de l'hypothèse de l'hétérogénéité, comment se fait-il que de nombreux psychologues cherchent encore à formuler une théorie unique, postulant un type homogène de concepts ? Les raisons de soutenir l'hypothèse de l'hétérogénéité sont claires, mais qu'en est-il des raisons qui motivent ces psychologues ? Leur comportement est-il purement irrationnel ?

C'est une question délicate. Il est toutefois possible que ces psychologues soient guidés, sans doute implicitement, par des considérations de *parcimonie* et d'*unification*. Il y a de nombreuses manières d'explicitier le principe de parcimonie. L'une d'entre elles est la suivante : toutes choses étant égales, une théorie qui postule moins d'entités est préférable à une théorie qui postule davantage d'entités. Le principe d'unification est le suivant (Friedman 1974, Kitcher 1981) : toutes choses étant égales, une théorie qui explique un ensemble donné de phénomènes est préférable à plusieurs théories qui expliquent ces mêmes phénomènes. Quoique ces deux principes soient normatifs, ils décrivent aussi l'activité de nombreux chercheurs dans de nombreux domaines scientifiques. En effet, nombre de scientifiques admettent ces deux principes, préférant les théories parcimonieuses et cherchant à unifier les phénomènes sous des théories générales.

Si les psychologues qui travaillent sur les concepts et les fonctions cognitives supérieures admettent ces deux principes, ils devraient être réticents à rejeter l'hypothèse EN au profit de l'hypothèse de l'hétérogénéité. Admettre l'hypothèse EN, c'est soutenir qu'un type homogène de concepts peut expliquer l'ensemble ou la plupart des phénomènes qui caractérisent les processus cognitifs supérieurs. Au contraire, l'hypothèse de l'hétérogénéité

propose de postuler plusieurs types très différents de concepts, au nombre desquels il faut au moins compter les prototypes, les exemplaires et les théories. Le principe de parcimonie favorise donc l'hypothèse EN au dépens de l'hypothèse de l'hétérogénéité. C'est aussi le cas du principe d'unification. Car l'hypothèse de l'hétérogénéité conduit à admettre différentes théories pour expliquer différents phénomènes (voir les chapitres 5 et 6) : la théorie des prototypes et de leur utilisation dans les processus cognitifs supérieurs explique certains phénomènes, la théorie des exemplaires et de leur utilisation dans les processus cognitifs supérieurs en expliquent d'autres etc.

Supposons donc que les psychologues préfèrent l'hypothèse EN parce qu'ils admettent le principe de parcimonie et le principe d'unification. Cela pose la question suivante : est-il erroné d'admettre ces deux principes ? La question est de toute évidence complexe et ne saurait être traitée pleinement dans l'espace de cette conclusion. Notons toutefois qu'il n'est pas question de rejeter entièrement ces deux principes. Dans certaines sciences, par exemple en physique, ils semblent jouer un rôle crucial dans l'évaluation des théories scientifiques. Par exemple, le principe de parcimonie semble important en physique des particules, puisque les physiciens cherchent à réduire le nombre de particules fondamentales. Le principe d'unification caractérise aussi certaines étapes du progrès scientifique : l'une des supériorités de la théorie de Newton résulte ainsi de sa capacité à rendre compte d'un large nombre de phénomènes empiriques (Kitcher 1981, voir aussi la critique des théories de la sélection des groupes par Williams).

On peut cependant douter de la validité universelle de ces deux principes. En effet, leur validité semble dépendre de la nature des problèmes étudiés. Je propose que le statut normatif de l'unification et de la parcimonie ne résulte pas de la nature de l'explication scientifique *per se*, mais de la nature des problèmes étudiés (voir l'approche localiste du principe de parcimonie dans Sober 1994). Dans certains contextes, ces deux principes sont légitimes, mais sans doute pas dans tous. Par exemple, le principe de parcimonie ne semble pas s'appliquer automatiquement en biologie moléculaire. Il est bien connu que les gènes manifestent la propriété de redondance. Plusieurs séquences d'ADN codent pour la même protéine. Cela protège la cellule contre les erreurs de transcription d'une séquence particulière. Pareillement, le principe d'unification ne semble pas s'appliquer automatiquement en biologie évolutionniste. Il est probable qu'il faille différentes théories

pour expliquer différents résultats de l'évolution (Mitchell 2003). Il est peut-être nuisible de chercher à étendre une unique approche à l'ensemble des phénomènes biologiques.<sup>5</sup>

Certains phénomènes étudiés en psychologie cognitive ne se prêtent guère à des explications scientifiques fondées sur ces deux principes. La cognition est en partie un produit de l'évolution. Différents types d'états mentaux, différents processus cognitifs ont peut-être été sélectionnés à différentes étapes de notre évolution. Par ailleurs, on peut supposer que nos processus cognitifs sont adaptés, par évolution ou par apprentissage, à des environnements et à des tâches spécifiques (voir Gigerenzer et al. 1999, Todd 2000). Cela implique que nous possédons de nombreux processus cognitifs, adaptés à des contextes spécifiques (voir la notion de boîte à outils développée par Gigerenzer et al. 1999). En conséquence, il est plausible que le principe de parcimonie ne s'applique pas toujours en psychologie cognitive, voire qu'il s'applique rarement. Cela implique aussi que les phénomènes cognitifs dépendent de différents mécanismes cognitifs. En conséquence, il est plausible que le principe d'unification ne s'applique pas toujours en psychologie cognitive, voire qu'il s'applique rarement.

Il est donc possible que les psychologues des concepts et des fonctions cognitives supérieures adhèrent si fortement à l'hypothèse EN pour des raisons, les principes de parcimonie et d'unification, qui, tout en étant pertinentes dans certains domaines scientifiques, ne le sont pas nécessairement en psychologie. Je ne prétends certes pas avoir établi que c'est là la principale raison des psychologues ni avoir prouvé que ces deux principes ne peuvent en aucun cas s'appliquer en psychologie cognitive. Toutefois, les paragraphes précédents montrent que ce sont là deux affirmations plausibles.

#### *2.4. Conclusion*

Dans ce chapitre, j'ai présenté et discuté les conséquences critiques de ce travail. Le rejet de l'hypothèse EN implique que l'une des controverses les plus importantes en psychologie cognitive, à savoir la controverse sur la nature des concepts, n'avait pas lieu d'être. Il existe différents types de concepts, différentes théories sont donc requises. J'ai cherché à évaluer les conséquences qu'ont eu l'hypothèse EN et cette controverse sur le travail empirique des psychologues et j'ai aussi discuté les raisons qui peuvent expliquer l'attachement des psychologues à l'hypothèse EN.

---

<sup>5</sup> La complexité des phénomènes biologiques est sans doute aussi incompatible avec le principe d'unification (Mitchell 2003).

### 3. De nouveaux thèmes de recherche

La seconde section est essentiellement critique : elle tire les leçons négatives du rejet de l'hypothèse EN. Cependant, l'hypothèse de l'hétérogénéité a aussi des conséquences positives. De nouveaux thèmes de recherche, aussi bien en philosophie qu'en psychologie cognitive, deviennent en effet saillants lorsque l'on accepte cette hypothèse. A plusieurs reprises au cours de ce travail, j'ai insisté sur ce point, déplorant ça et là que les psychologues n'aient pas davantage travaillé sur tel ou tel sujet. Dans cette section, je résume ces nouveaux thèmes de recherche.

#### *3.1. Philosophie des concepts : de la philosophie de l'esprit à la philosophie de la psychologie*

Dans l'introduction et le premier chapitre de ce travail, j'ai rappelé que les concepts étaient habituellement un thème de la philosophie de l'esprit. J'ai écarté les problèmes posés habituellement dans ce domaine. L'une des ambitions de ce travail était en effet de redessiner la carte des problèmes liés aux concepts, en montrant que la psychologie des concepts et des fonctions cognitives supérieures posent des problèmes qui, malgré leur intérêt philosophique, ont été jusqu'à présent négligés. La validité de l'hypothèse EN est l'un de ces problèmes.<sup>6</sup> Ainsi, je propose que les concepts deviennent un thème de la philosophie de la psychologie plutôt que de la philosophie de l'esprit.

Résumons brièvement les principales différences entre le traitement des concepts en philosophie de l'esprit et le traitement des concepts en philosophie de la psychologie.

- En philosophie, la notion de concept est habituellement utilisée pour désigner les composants de nos attitudes propositionnelles. En d'autres termes, les philosophes supposent que les croyances et les désirs sont des états mentaux complexes, qui, d'une manière ou d'une autre, sont composés par les concepts (voir le chapitre 1, section 1.3.). Je propose au contraire de prendre pour point de départ la manière dont les psychologues caractérisent les concepts : les concepts sont les ensembles de données à propos de classes d'entités qui sont stockés en mémoire à long terme et que nous utilisons par défaut dans nos processus cognitifs supérieurs.
- Les questions posées en philosophie de l'esprit sont différentes des questions que l'on pourrait poser en philosophie de la psychologie. Un petit nombre de questions sont débattues en philosophie de l'esprit. En particulier, la plupart des arguments avancés en philosophie de l'esprit à propos des concepts portent sur les deux questions suivantes :

---

<sup>6</sup> Le nouvel empirisme (Barsalou 1999, Prinz 2002) est un autre de ces thèmes. Voir aussi Margolis 1994, 1995.

- Comment individuer les concepts ? Faut-il prendre en compte les inférences dans lesquelles les concepts figurent ? Et si oui, lesquelles ?
- D'où viennent les propriétés sémantiques des concepts ?

Par ailleurs, le thème des concepts en philosophie de l'esprit est souvent associé à des questions épistémologiques, en particulier à des questions sur les fondements de la justification (voir l'introduction). Les questions posées en philosophie de la psychologie résultent de la psychologie des concepts et des fonctions cognitives supérieures. Elles visent à éclaircir certaines notions liées aux concepts, par exemple la notion de théorie (Margolis 1995), à évaluer la solidité des arguments avancés en psychologie (par exemple, Margolis 1994) et à développer théoriquement certaines notions ou positions (par exemple, Prinz 2002).

- Les méthodes utilisées dans ces deux branches sont aussi différentes. Les philosophes de l'esprit font appel à leurs intuitions. Ainsi, pour expliquer les propriétés sémantiques des concepts, les philosophes font appel à leurs intuitions sur la référence de ces représentations mentales. Pour individuer les concepts, ils font appel à leurs intuitions sur les conditions d'identité des concepts et des attitudes propositionnelles, par exemple des croyances (voir le principe de Frege dans Peacocke 1992). Au contraire, en philosophie de la psychologie, la notion de concept est vue comme une notion théorique. En conséquence, nos intuitions n'ont guère de pertinence. Par ailleurs, il devient nécessaire de prendre en compte les données empiriques récentes sur les ensembles de données que nous utilisons par défaut dans nos processus cognitifs supérieurs et sur ces processus cognitifs eux-mêmes.

Bien que l'objet de ce travail ne soit pas de critiquer les problèmes posés par les philosophes de l'esprit à propos des concepts et les méthodes qu'ils utilisent, j'espère que ce travail montre, par l'exemple, le potentiel de la philosophie de la psychologie. J'aimerais que ce travail contribue, avec d'autres travaux (Margolis 1994, 1995, Prinz 2002), à une réorientation de la philosophie des concepts.

### 3.2. *Questions non résolues*

J'ai soutenu que nous possédions au moins trois types fort différents de concepts, à savoir des prototypes, des exemplaires et des théories. J'ai décrit les thèses centrales des paradigmes qui postulent ces types de représentations mentales (voir le chapitre 3). Toutefois, j'ai souligné

que ces paradigmes ne pouvaient être identifiés à des théories (voir le chapitre 3, section 1.). Ils décrivent seulement les propriétés principales des concepts. Plusieurs théories peuvent articuler un même paradigme. Il existe par exemple plusieurs théories des exemplaires. Elles spécifient différemment la nature des exemplaires, leurs propriétés sémantiques, la manière dont ils sont formés, la manière dont ils sont employés dans nos processus cognitifs supérieurs. De toute évidence, les psychologues doivent élaborer des théories précises de ces types de représentations mentales et les évaluer au moyen de données empiriques.

Par ailleurs, j'ai reconnu que, outre les prototypes, les exemplaires et les théories, nous possédions peut-être d'autres types de concepts (voir le chapitre 3). L'hypothèse de l'hétérogénéité rend donc saillantes les questions suivantes : combien de types de concepts convient-il de postuler ? Quels sont-ils ? Quelles sont leurs propriétés ? C'est une question empirique à laquelle on ne peut répondre *a priori*.

Notons que pour introduire un nouveau type de concepts, deux conditions doivent être remplies.

1. Il faut montrer qu'il est *nécessaire* de postuler ce type de concepts *pour expliquer* certaines propriétés des processus cognitifs ; pour ce faire, il faut montrer que les connaissances que ce type de concepts est supposé encoder sont utilisées par défaut par nos processus cognitifs – au moins, dans certains contextes. Montrer que nous sommes en mesure d'utiliser ces connaissances ne suffit pas à établir qu'elles ont pour fonction d'être utilisées par défaut dans nos processus cognitifs (voir le critère de démarcation dans le chapitre 1, section 1.2.3.).
2. Il faut aussi montrer que ce type de concepts est fort différent des autres types de concepts, et non un type particulier de prototypes, d'exemplaires ou de théories.

J'ai en particulier proposé que les concepts incarnés qui ont été postulés par Barsalou et Prinz (voir le chapitre 3, section 3.3.) formaient sans doute un autre type de concepts, différent des trois types de concepts examinés dans ce travail. A la différence de Prinz et Barsalou, je crois toutefois qu'il est vain de chercher à montrer que *tous* les concepts sont des symboles perceptuels, en forçant toutes nos représentations mentales dans le cadre étroit de l'empirisme. L'hypothèse de l'hétérogénéité suggère plutôt de chercher à déterminer quels types de phénomènes empiriques montrent que, dans certains contextes, nous utilisons par défaut ce type de connaissances.

### *3.3. Nouveaux thèmes en psychologie : la théorie pluraliste de la cognition*

L'hypothèse de l'hétérogénéité est liée à la théorie pluraliste de la cognition. Selon la théorie pluraliste, une même compétence cognitive peut reposer sur plusieurs processus cognitifs. Ces processus cognitifs peuvent faire appel à des ensembles de connaissances fort différents les uns des autres (voir le schéma 4.5. dans le chapitre 4). Plusieurs programmes de recherche ont développé des idées semblables. Le programme de recherche « Heuristiques rapides et frugales », qui a été développé au Centre pour le Comportement et la Cognition Adaptative (Institut Max-Planck pour le Développement Humain, voir Gigerenzer et al. 1999), propose que l'esprit contient de nombreuses heuristiques propres à des contextes particuliers. Par ailleurs, les théories duelles de la cognition supposent qu'une même compétence, par exemple le raisonnement déductif, repose souvent sur deux processus (voir le chapitre 4, section 3.). Malgré les différences entre la théorie pluraliste de la cognition, les théories duelles de la cognition et le programme de recherche Heuristiques rapides et frugales, ils affirment tous l'existence de plusieurs processus cognitifs pour une même compétence cognitive. Par exemple, ils s'accordent tous pour dire qu'il existe plus d'un processus cognitif de catégorisation.

Même si un nombre croissant de psychologues semblent maintenant embrasser des idées plus ou moins semblables, la théorie pluraliste de la cognition n'en a pas moins été pendant longtemps marginale. Cette négligence est sans doute partiellement l'effet de l'hypothèse EN. Alors que l'hypothèse EN s'accorde naturellement avec l'idée qu'une compétence cognitive repose sur un unique processus cognitif (sans l'impliquer toutefois), le rejet de l'hypothèse EN rend saillante la théorie pluraliste de la cognition (sans l'impliquer non plus). Supposons en effet que les concepts forment un type homogène ; autrement dit, admettons l'hypothèse EN. Tous les concepts ont en conséquence des propriétés fonctionnelles semblables. Par exemple, si tous les concepts sont des théories, les processus cognitifs supérieurs reposent tous sur des règles (voir le chapitre 4, section 2.4.). Certes, cela n'implique nullement l'existence d'un unique processus conceptuel pour une compétence donnée, car il pourrait exister plusieurs processus fondés sur des règles. Par exemple, il pourrait sans doute exister plusieurs processus de catégorisation fondés sur des règles. Cependant, comme l'hypothèse EN postule une classe homogène de concepts, il semble naturel de postuler l'existence d'un unique processus conceptuel, par exemple un unique processus pour catégoriser. Admettons, au contraire, l'hypothèse de l'hétérogénéité. Elle n'implique pas l'existence de plusieurs processus cognitifs. Un unique processus pourrait

prendre pour inputs ces différents types de représentations (voir le chapitre 6, pour une proposition analogue). Cependant, elle rend plausible l'existence de plusieurs processus.

Adopter l'hypothèse de l'hétérogénéité conduit donc à examiner la théorie pluraliste de la cognition et les problèmes qu'elle pose. En conséquence, le rejet de l'hypothèse EN rend saillants plusieurs thèmes de recherche qui n'ont pas reçu l'attention qu'ils méritent. Nous avons vu à plusieurs reprises que plusieurs questions étaient encore sans réponse.

Un premier type de questions concerne *l'articulation de nos processus conceptuels*. Par exemple, si nous avons plusieurs processus pour catégoriser, il est nécessaire de comprendre comment ces processus sont articulés les uns aux autres. On peut imaginer plusieurs options :

1/ Ces processus sont déclenchés en parallèle. Par exemple, chaque fois que nous catégorisons, nos différents processus pourraient être déclenchés.

2/ Ils sont au contraire déclenchés dans des contextes différents. Un processus pourrait être déclenché dans un type particulier de contextes, tandis qu'un autre processus pourrait être déclenché dans un autre type de contextes.

De nombreuses hypothèses intermédiaires sont envisageables :

3/ Le déclenchement des processus cognitifs est sensible au contexte, mais certains contextes déclenchent tous les processus.

4/ Il se peut aussi qu'un processus soit toujours déclenché, tandis que le déclenchement des autres processus soit sensible au contexte (les avocats des théories duelles des processus cognitifs soutiennent souvent cette hypothèse, voir le chapitre 3, section 3.). Nos processus pourraient donc fonctionner seulement à l'occasion en parallèle.

Un second type de questions concerne *la nature du déclenchement de ces processus*.

1/ Ces processus sont-ils déclenchés automatiquement ou bien doivent-ils être déclenchés volontairement ?

2/ Il se peut qu'un processus soit automatiquement déclenché, tandis que les autres sont déclenchés volontairement. Notons que cela ne revient pas à dire que l'un des processus est toujours déclenché, car le déclenchement d'un processus peut être à la fois spécifique à certains contextes et automatique.

Des hypothèses plus complexes sont aussi envisageables :

3/ Dans certains contextes, un processus est déclenché automatiquement, tandis qu'il doit être déclenché intentionnellement dans d'autres contextes.

Un troisième type de questions concerne *l'intégration des outputs de ces processus*. A supposer que, au moins dans certaines circonstances, nos processus soient déclenchés en parallèle, comment les outputs de ces processus sont-ils intégrés ?

- 1/ Les outputs d'un type de processus, par exemple, des processus fondés sur des règles, prévalent-ils sur les outputs des autres processus ?
- 2/ Le processus le plus rapide prévaut-il sur les autres processus ?
- 3/ Les outputs sont-ils intégrés ? Et comment ?

Un quatrième type de questions concerne *la sensibilité au contexte de nos processus cognitifs*. Si l'on suppose que certains processus sont sensibles au contexte, il faut expliquer la nature de cette sensibilité au contexte. Plusieurs questions se posent :

- 1/ Dans quel contexte un processus est-il déclenché ? Par exemple, si nous avons différents processus cognitifs pour catégoriser, qui emploient des types de concepts différents, et si ces processus cognitifs sont déclenchés dans des types de contextes différents, quel contexte déclenche quel processus cognitif ? Répondre à cette question implique d'étudier les propriétés des environnements dans lesquels nous utilisons nos processus cognitifs. Il faut caractériser les propriétés des contextes qui déclenchent tel processus cognitif plutôt que tel autre.
- 2/ Et pourquoi ? Il est possible que nos processus cognitifs soient, en règle générale, adaptés aux environnements dans lesquels ils sont utilisés (voir la notion de rationalité écologique dans Gigerenzer et al. 1999). Autrement dit, nos processus cognitifs sont déclenchés pour certaines fins. En règle générale, ils nous permettent d'atteindre ces fins. Et ceci grâce à une adéquation entre leur structure et les propriétés des environnements dans lesquels ils sont utilisés (voir l'heuristique de reconnaissance dans Goldstein et Gigerenzer 2002). On peut supposer que c'est aussi le cas de l'utilisation de théories, d'exemplaires et de prototypes. Dans ce cas, il faudrait examiner quelles propriétés des environnements et des tâches cognitives rendent l'utilisation des prototypes, l'utilisation d'exemplaires et l'utilisation de théories efficaces. En d'autres termes, il faudrait étudier la rationalité écologique de l'utilisation d'exemplaires, de l'utilisation de prototypes et de l'utilisation de théories.

Il existe déjà un certain nombre de *données empiriques* qui sont pertinentes pour répondre à ces questions (Allen et Brooks 1991, Smith et Sloman 1994, Sloman 1996, 1998, Smith et al. 1998, Gigerenzer et al. 1999, Stanovich et West 2000, 2003). Cependant, ces

données sont encore trop peu nombreuses. Par ailleurs, les psychologues ont surtout cherché à montrer qu'il existe plusieurs processus (Allen et Brooks 1991, Smith et Sloman 1994, Haidt 2001, Greene et al. 2001...). En conséquence, les données empiriques pertinentes n'ont pas été obtenues pour chercher à distinguer différentes réponses possibles aux questions posées ci-dessus. Enfin, il est sans doute temps de formuler des théories précises sur l'articulation des différents processus cognitifs dont dépend une compétence cognitive.

#### *3.4. Nouveaux thèmes en psychologie : la question de la coordination et l'évolution des concepts*

Admettons maintenant qu'il existe plusieurs types de concepts qui appartiennent à des types différents de représentations mentales – entre autres des prototypes, des exemplaires et des théories. Nous avons vu que les concepts coréférentiels sont très souvent *coordonnés* (voir le chapitre 8, section 3.2.). L'activation d'un concept active les autres concepts coréférentiels : par exemple, catégoriser un chien au moyen du prototype de chien permet de faire des inférences à propos des chiens sur la base de notre concept théorique de chien. Il existe plusieurs manières de concevoir la coordination des concepts coréférentiels (voir le chapitre 8, section 3.2.3.). Il n'existe par ailleurs guère de données empiriques pertinentes. Si l'on admet l'hypothèse de l'hétérogénéité, il est important de formuler des thèses précises sur la coordination et de trouver les moyens expérimentaux pour tester ces théories et distinguer les différentes théories possibles.

Par exemple, si l'on suppose que les différents concepts coréférentiels sont coordonnés parce qu'ils sont tous coordonnés à la même représentation linguistique (voir le chapitre 8, schéma 8.2.), la coordination des concepts devrait être affectée chez certains patients qui souffrent de troubles du langage. Cela n'est pas le cas si l'on suppose que les concepts sont directement coordonnés les uns aux autres (voir le chapitre 8, schéma 8.1.). Par ailleurs, différents concepts peuvent être coordonnés différemment. L'origine, le développement et la stabilité de la coordination devraient aussi être des objets d'étude.

La cognition, comme de nombreux autres aspects de notre phénotype, est, au moins pour partie, le produit de l'évolution. Admettons ici l'utilité des principes évolutionnistes en psychologie (Tooby et Cosmides 1992). L'hypothèse de l'hétérogénéité pose alors la question suivante : *comment expliquer l'évolution de plusieurs types de concepts ?* Pourquoi encoder de cette manière les informations à propos des classes d'entités ? De nombreuses hypothèses évolutionnistes sont concevables. On pourrait, par exemple, proposer l'explication suivante.

L'évolution se produit par modification des phénotypes préexistants. Il est possible que prototypes, exemplaires et théories correspondent à différents moments de l'évolution de la cognition. De toute évidence, d'autres hypothèses sont possibles. Il faut spécifier précisément ces hypothèses évolutionnistes et déterminer des manières de les tester – par exemple, au moyen de comparaisons cross-spécifiques ou au moyen de prédictions psychologiques.

### *3.5. Conclusion*

Dans cette section, j'ai décrit les apports positifs de ce travail. Plutôt que d'être un thème en philosophie de l'esprit, les concepts peuvent être traités en philosophie de la psychologie. Cette approche permet de poser de nouveaux problèmes largement négligés par les philosophes. Par ailleurs, en psychologie, de nouveaux thèmes de recherche sont rendus saillants par le rejet de l'hypothèse EN. C'est en particulier le cas de l'articulation des processus cognitifs et de la coordination des concepts.

## **4. L'avenir des concepts**

### *4.1. Le problème*

On trouve, dans de nombreuses disciplines scientifiques ainsi qu'en philosophie des sciences, des arguments semblables à l'argument proposé dans ce travail. Par exemple, en philosophie de la biologie, Marc Ereshefsky (1998) soutient que la notion d'espèce réfère à une classe de taxa extrêmement hétérogène. Il en tire deux conclusions. Tout d'abord, la notion d'espèce n'a guère de rôle à jouer en biologie et devrait donc être éliminée. Il conclut aussi que l'existence de la catégorie d'espèce est douteuse (Ereshefsky 1998, 118). En philosophie de la psychologie, Paul Griffith (1997) a pareillement soutenu que la notion d'émotion réfère à une classe extrêmement hétérogène. Il en conclut que le terme « émotion » n'a guère d'importance pour la psychologie (contrairement à d'autres notions, comme la notion de programme affectif développée par Ekman – « affect program »), même si elle peut être importante dans d'autres domaines, par exemple dans le domaine des relations sociales (Griffith 2003). On trouve des arguments semblables en psychologie. Par exemple, de nombreux psychologues ont soutenu que la notion d'apprentissage faisait en fait référence à des mécanismes cognitifs qui ont fort peu en commun (Gallistel 1999). On peut alors à juste titre douter de l'utilité de la notion d'apprentissage en psychologie.

Dans tous ces cas, on infère *l'inutilité d'une notion scientifique de l'hétérogénéité* de la classe dénotée. Cela suggère d'éliminer la notion scientifique X et de la remplacer par

d'autres notions plus adéquates, Y, Z... On peut alors vouloir inférer que les x n'existent pas. Il est tentant de faire appel à ce type d'arguments dans le présent contexte. D'où les questions : Quelle est l'utilité de la notion de concept quand on rejette l'hypothèse EN au profit de l'hypothèse de l'hétérogénéité ? Quelle est la validité des arguments éliminativistes proposés ci-dessus ? Dans quelles conditions sont-ils valides ? Impliquent-ils que nous n'avons pas de concepts ? Dans cette section, je m'efforce de répondre à ces questions.

#### 4.2. *Un problème mineur*

Auparavant, insistons encore une fois sur le message central de ce travail. J'ai cherché à montrer que les concepts ne forment pas une espèce naturelle (section 1). Autrement dit, il n'est pas légitime de prétendre caractériser la classe entière des ensembles de représentations utilisées par défaut dans nos processus cognitifs à partir des propriétés possédées par certains concepts. Au contraire, d'autres classes, comme les classes des prototypes, des exemplaires et des théories, sont probablement les classes adéquates pour décrire ces représentations. Différentes théories sont nécessaires pour caractériser ces différentes classes. Cela implique que la classe des concepts ne joue pas le rôle crucial dans les théories de la cognition supérieure qu'elle est censée jouer selon de nombreux psychologues.

*Cette thèse est indépendante de l'avenir de la notion de concept.* Qu'il faille l'éliminer ou qu'il faille la conserver, il n'en reste pas moins que les représentations qui sont utilisées par défaut dans nos processus cognitifs supérieurs ne forment pas une espèce naturelle. En particulier, même si l'on conserve le terme « concept » (voir la discussion ci-dessous), son importance serait considérablement réduite. Il ne désignerait pas l'une de ces classes importantes qui donnent lieu à des généralisations. Il ne figurerait pas dans des généralisations empiriques. Il serait remplacé dans ces rôles par d'autres termes, comme « prototype », « exemplaire » et « théorie ». Par ailleurs, le rejet de l'hypothèse EN rend saillantes de nouvelles questions et devrait réorienter en partie la psychologie des concepts (section 3.). La pertinence de ces nouveaux objets de recherche est aussi indépendante de l'avenir de la notion de concept. L'élimination de cette notion est donc un problème *mineur*.

#### 4.3. *Éliminer ou conserver ?*

Commençons par la nécessité de conserver ou d'éliminer une notion scientifique. La question ontologique (les X existent-ils ?) sera examinée par la suite. Il existe plusieurs types

d'arguments éliminativistes. Par exemple, on a souvent utilisé les types d'arguments suivants :

- *Argument de la confusion* : il est impossible d'explicitier la notion scientifique X de manière cohérente ou non circulaire (par exemple, Quine 1951 à propos des notions d'analyticité et de signification, Okasha 2003 à propos de la notion de sélection des clades). Donc, la notion X doit être éliminée.
- *Argument des fausses présuppositions* : la notion X est définie par des énoncés empiriques faux (par exemple, Churchland 1981 à propos de la notion de croyance). Donc, la notion X doit être éliminée.
- *Argument de l'hétérogénéité* : la notion X réfère à une classe d'entités très hétérogènes (par exemple, Kelly 1995 à propos de la notion de chasseur-cueilleur, Griffith 1997). Donc, la notion X doit être éliminée.

L'argument de la nécessité explicative fournit la prémisse pour un argument de ce troisième type. Concentrons-nous sur cet argument. La prémisse et la conclusion de cet argument sont les suivantes :

- P1 : La notion de concept ne réfère pas à une espèce naturelle.
- Cl : Il faut éliminer la notion de concept de la psychologie scientifique.

La prémisse P1 est la conclusion de l'argument de la nécessité explicative (ANE) et est impliquée par l'hypothèse de l'hétérogénéité (HH). La conclusion Cl est la thèse éliminativiste contre la notion de concept. Notons qu'il s'agit d'éliminer la notion de concept *de la psychologie scientifique*. Ce qui est en jeu, ce n'est pas l'élimination intégrale du terme « concept », par exemple du langage ordinaire, ni même l'élimination de ce terme de toutes les disciplines scientifiques (histoire des sciences, anthropologie...). La question porte sur la légitimité d'un terme théorique particulier, « concept », dans une discipline scientifique particulière, la psychologie cognitive.

Comment inférer la conclusion Cl à partir de P1 ? Il me semble qu'il y a deux options :

- P2 : Parmi les termes généraux, seuls les termes qui réfèrent à des espèces naturelles sont légitimes dans une discipline scientifique.

P2' : Il faut éliminer les termes généraux dont on a supposé à tort qu'il faisait référence à des espèces naturelles.

La prémisse P2' est de toute évidence plus faible que la prémisse P2 : P2' n'implique pas que seuls les termes généraux qui font référence à des espèces naturelles sont légitimes. P2' est compatible avec l'emploi légitime dans une science donnée de termes généraux qui ne réfèrent pas à des espèces naturelles. Mais P2' soutient que si un terme général X est supposé à tort référer à une espèce naturelle, il convient alors de l'éliminer.

L'une de ces deux prémisses est-elle justifiée ? Avant même de répondre à cette question, il faut d'abord examiner la question suivante : comment décider si l'une de ces deux prémisses est justifiée ? Comment déterminer les principes normatifs qui gouvernent le bon usage du vocabulaire scientifique ? Il est aujourd'hui largement admis que les propositions concernant les principes normatifs qui gouvernent l'activité scientifique doivent être compatibles avec l'activité scientifique elle-même : on suppose par là qu'en règle générale, l'activité scientifique est normativement correcte.<sup>7</sup> Etant donné ce principe méthodologique, il importe de savoir comment les scientifiques ont utilisé les termes généraux. En particulier, il faut déterminer si l'utilisation scientifique des termes généraux se conforme à P2 et à P2'. C'est-à-dire, il faut chercher à déterminer si les scientifiques utilisent des termes généraux qui ne réfèrent pas à des espèces naturelles (pour évaluer P2) et s'ils éliminent les termes généraux quand ils découvrent qu'ils ne réfèrent pas à des espèces naturelles (pour évaluer P2').

Notons que seuls les cas où la catégorie dénotée est extrêmement hétérogène sont pertinents dans le présent contexte.<sup>8</sup> La notion d'insecte fait référence à des espèces fort différentes les unes des autres. Mais, comme les insectes forment un taxon phylogénétique, de nombreuses propriétés sont non accidentellement vraies de tous les insectes : les insectes forment donc une espèce naturelle (voir le chapitre 2, section 1.3.4.). Pareillement, la classe des infections est certainement hétérogène. Mais on peut formuler des généralisations à propos de cette classe. En particulier, les approches évolutionnistes de la contamination

---

<sup>7</sup> Un tel présupposé peut être défendu de plusieurs manières. On pourrait, par exemple, mettre en avant les remarquables succès des sciences et conclure, par une inférence à la meilleure explication, que l'activité scientifique est probablement normativement correcte.

<sup>8</sup> Il y a d'autres raisons de soutenir qu'une classe n'est pas une espèce naturelle (voir le chapitre 2, section 1.3.4.). Par exemple, la notion de corps supralunaires, essentielle dans la physique aristotélicienne, n'est pas une espèce naturelle, parce qu'il n'y a pas de propriétés spécifiques pour ces corps. Il n'y a donc aucune raison de les distinguer des corps sublunaires.

permettent de formuler des généralisations nomologiques concernant la classe entière des infections (Ewald 1996).<sup>9</sup>

Soulignons par ailleurs qu'il est loin d'être certain que l'usage des termes généraux par les scientifiques puisse être caractérisé par des principes normatifs. En effet, il est possible que des considérations d'ordre social, polémique, politique etc., gouvernent en dernière analyse l'usage du vocabulaire scientifique (Stich 1996). Cela est en particulier le cas pour l'élimination ou le maintien des termes scientifiques. Il est par exemple possible que l'élimination d'un terme scientifique soit déterminée par des considérations de politique scientifique : éliminer un terme scientifique au profit de nouveaux termes peut être un moyen d'attirer l'attention de la communauté scientifique, maintenir un terme scientifique peut être un moyen de faire accepter une nouvelle théorie.

Ignorer l'importance de ces considérations dans la pratique scientifique conduit à une image naïve et appauvrie des sciences. Notons toutefois la distinction suivante. Il faut distinguer le cas où des considérations très variées, *y compris* des considérations sociales ou politiques, des considérations de carrière, des facteurs individuels etc., déterminent l'utilisation des termes scientifiques du cas où *seules* des considérations de ce type la déterminent. Supposons que la première possibilité soit le cas. Admettons aussi que les principes méthodologiques doivent être compatibles avec les pratiques scientifiques. Cela implique-t-il qu'il soit vain de chercher des principes normatifs ? Pas nécessairement. Car l'on pourrait chercher des principes *ceteris paribus*. Par exemple, on pourrait établir que *ceteris paribus*, il faut éliminer les termes scientifiques qui ne dénotent pas des espèces naturelles : si aucune autre considération n'est pertinente, alors il faut éliminer ces termes. On pourrait répondre à cette question par une analyse des cas pertinents dans l'histoire des sciences. En contrôlant les considérations personnelles, sociales etc., on pourrait voir apparaître des principes *ceteris paribus*.

Peut-on inférer des principes gouvernant l'élimination des notions scientifiques à partir de l'usage du vocabulaire scientifique, lorsque ces notions dénotent des classes qui ne forment pas des espèces naturelles ? Et si oui, quels sont-ils ? En particulier, viennent-ils à l'appui de P2 et P2' ? Faute d'une étude complète sur l'usage des termes scientifiques, je ne peux apporter de réponse décisive à ces questions. *C'est une question ouverte.*

---

<sup>9</sup> Soulignons aussi que la notion d'infection n'est pas seulement une notion scientifique.

#### 4.4. Avons-nous des concepts ?

Les arguments pour l'élimination d'une notion scientifique sont souvent liés à des thèses ontologiques (Churchland 1981, Stich 1983, 1996). Par exemple, Ereshefsky (1998) infère de l'hétérogénéité des taxa nommés « espèce » que la catégorie des espèces n'existe pas. Supposons que nous devions éliminer le terme « concept ». *Devrait-on conclure (i) que nous n'avons pas de concept ou (ii) que la catégorie des concepts n'existe pas ?* La réponse à ces deux questions est de toute évidence *négative*. Considérons, par exemple, le cas des corps supralunaires. Dans la physique aristotélicienne, les corps sublunaires sont supposés former une espèce naturelle. Ils sont supposés avoir en commun de nombreuses propriétés : par exemple, ils sont supposés être formés de la même matière, leur orbite est supposée être parfaite etc. Depuis la révolution galiléenne, nous ne tenons plus les corps supralunaires pour une espèce naturelle (voir la note 8). Par ailleurs, le terme « supralunaire » ne fait plus partie de la physique depuis plusieurs siècles. Mais nous n'avons de toute évidence pas conclu que les corps supralunaires n'existaient pas. La Lune, les planètes du système solaire et les étoiles sont bien réelles. Bien plus, la classe des corps supralunaires est aussi bien réelle – bien des classes ne sont pas des espèces naturelles. Même s'il convenait d'éliminer le terme « concept » de la psychologie scientifique – et nous avons vu que la question ne peut être simplement tranchée, on ne pourrait conclure à la non-existence des concepts. Les concepts existent bien.

### 5. Conclusion

Les concepts ne forment pas une espèce naturelle. C'est là le message principal de ce travail : la classe des concepts n'est pas pertinente pour formuler des généralisations à propos des représentations mentales qui sont stockées en mémoire à long terme et qui sont utilisées par défaut dans nos processus cognitifs. Elle ne joue donc pas le rôle qu'elle était supposée jouer dans les théories de la cognition. Ce rôle est au contraire tenu par les notions de prototypes, d'exemplaires, de théories et sans doute par d'autres notions. La notion de concept perd donc l'essentiel de sa pertinence. La controverse sur la nature des concepts, qui provient peut-être d'un désir d'unification théorique ou d'un souci de parcimonie, n'avait donc guère lieu d'être : il est vain de chercher à caractériser les propriétés qui seraient vraies de tous les concepts. En revanche, de nombreuses questions deviennent pressantes une fois que l'on rejette l'hypothèse EN : comment caractériser précisément les prototypes, les théories, et les exemplaires ? Faut-il admettre d'autres types de concepts ? Comment les processus cognitifs

sont-ils articulés ? Comment nos concepts coréférentiels sont-ils coordonnés ? Le rejet de l'hypothèse EN, on le voit, pourrait conduire à une réorientation de la psychologie des concepts. Quel rôle reste-t-il pour la notion de concept ? Doit-on la conserver ? Il serait agréable de pouvoir apporter une réponse simple à cette question. Hélas, la nature des principes qui gouvernent le bon usage des termes scientifiques est obscure. Mais, au fond, cela importe peu : l'essentiel, c'est d'abandonner l'hypothèse EN et de réaliser que la notion de concept ne peut jouer le rôle qu'elle est traditionnellement censée jouer. Les concepts ne forment pas une espèce naturelle.

## Bibliographie (445 références)

---

- Abbott, B., 1997, A note on the nature of water, *Mind*, 106, 311-319.
- Abbott, B., 1999, Water = H<sub>2</sub>O, *Mind*, 108, 145-148.
- Ahn, W. K., 1998, Why are different features central for natural kinds and artifacts ? The role of causal status in determining feature weights, *Cognition*, 69, 135-178.
- Ahn, W. K., et Dennis, M. J., 2001, Dissociation between categorization and similarity judgement : differential effect of causal status on feature weights, dans U. Hahn et M. Ramscar, éd., 2001.
- Ahn, W., Kalish, Ch., Gelman, S. A., Medin, D. L., Luhmann, Ch., Atran, S., Coley, J. D., Shafto, P., 2001, Why essences are essential in the psychology of concepts, *Cognition*, 82, 1, 59-69.
- Allen, C., 1999, Animal concepts revisited : The use of self-monitoring as an empirical approach, *Erkenntnis*, 51, 1, 33-40.
- Allen, S. W., et Brooks, L. R., 1991, Specializing the operation of an explicit rule, *Journal of Experimental Psychology : General*, 120, 3-19.
- Anderson, J. R., 1990, The adaptive nature of human categorization, *Psychological Review*, 98, 409-429.
- Andler, D., 1990, La représentation des structures : un problème pour le connexionnisme, dans *La Machine perceptive, Actes des 12e journées francophones sur l'informatique*, Paris : EC2 éditeur.
- Aristote, 1997, *De l'Ame*, traduit par R. Bodeus, Paris : GF Flammarion.
- Armstrong, S. L., Gleitman, L. R., et Gleitman, H., 1983, What some concepts might not be, *Cognition*, 13, 263-308.
- Ashby, F. G., et Maddox, W. T., 1993, Relations between exemplar, prototype, and decision bound models of categorization, *Journal of Mathematical Psychology*, 37, 372-400.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A.U., et Waldron, E. M., 1998, A neuropsychological theory of multiple systems in category learning, *Psychological Review*, 105, 442-481.
- Atran, S., 1990, *Cognitive foundations of natural history : Towards an anthropology of science*, New York, NY : Cambridge University Press.

- Atran, S., 1994, Core domains versus scientific theories : Evidence from systematics et Itza-Maya folkbiology, dans L. A. Hirschfeld et S. A. Gelman, éd., 1994.
- Atran, S., 1999, Itzaj Maya folkbiological taxonomy : Cognitive universals and cultural particulars, dans D. L. Medin et S. Atran, éd., 1999.
- Au, T. K., et Romo, L. F., 1999, Mechanical causality in children's « folkbiology », dans D. L. Medin et S. Atran, éd., 1999.
- Aydede, M., 1998, Fodor on concepts and Frege puzzles, *Pacific Philosophical Quarterly*, 79, 289-294.
- Baillargeon, R., 1993, The object concept revisited : new directions in the investigation of infants' physical knowledge, dans C. Granrud, éd., 1993, *Visual perception and cognition dans infancy*, Hillsdale, NJ : Erlbaum.
- Baillargeon, R., Kotovsky, L., et Needham, A., 1995, The acquisition of physical knowledge in infancy, dans D. Sperber, D. Premack, et A. J. Premack, éd., *Causal cognition : A multidisciplinary debate*, New York : Oxford University Press.
- Barrett, H. C., 2001, On the functional origins of essentialism, *Mind and Society*, 3, 2, 1-30.
- Barsalou, L. W., 1983, Ad hoc categories, *Memory and cognition*, 10, 82-93.
- Barsalou, L. W., 1985, Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 11, 629-654.
- Barsalou, L. W., 1987, The instability of graded structure : Implications for the nature of concepts, dans U. Neisser, éd., 1987.
- Barsalou, L. W., 1990, On the indistinguishability of exemplar memory and abstraction in category representation, dans T. K. Srull et R. S. Wyer, éd., *Advances in social cognition, Vol. III, Content and process specificity in the effects of prior experiences*, Hillsdale, NJ : Erlbaum.
- Barsalou, L. W., 1999, Perceptual symbol systems, *Behavioral and Brain Sciences*, 22, 577-660.
- Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., et Ruppert, J. A., 2003a, Social embodiment, dans B. H. Ross, éd., *The Psychology of learning and motivation*, vol. 43, San Diego, CA : Academic Press.
- Barsalou, L. W., Simmons, W. K., Barbey, A., et Wilson, C. D., 2003b, Grounding conceptual knowledge in modality-specific systems, *Trends in Cognitive Sciences*, 7, 84-91.
- Barsalou, L. W., Solomon, K. O., et Wu, L. L., 1999, Perceptual simulation in conceptual tasks, dans M. K. Hiraga, C. Sinha, et S. Wilcox, éd., *Cultural, typological, and psychological*

- in cognitive linguistics : The Proceedings of the 4th conference of the International Cognitive Linguistics Association*, Vol. 3, Amsterdam : John Benjamins.
- Berkeley, G., 1710/1998, *A treatise concerning the principles of human knowledge*, édité par J. Dancy, Oxford : Oxford University Press.
- Berretty, P. M., Todd, P. M., et Martignon, L., 1999, Categorization by elimination, dans G. Gigerenzer et P. M. Todd, éd..
- Bishop, M. A., et Stich, S. P., 1998, The flight to reference, or how not to make progress in the philosophy of science, *Philosophy of Science*, 65, 33-49.
- Block, N., 1986, Advertisement for a semantics for psychology, dans P. A. French, T. E. Uehling Jr., et H. K. Wettstein, éd., *Midwest studies in philosophy X: Studies in the philosophy of mind*, Minneapolis : University of Minnesota Press.
- Block, N., 1993, Holism, hyper-analyticity and hyper-compositionality, *Mind and Language*, 8, 1-26.
- Bloom, P., 1996, Intention, history, and artifact concepts, *Cognition*, 60, 1-29.
- Bloom, P., 2000, *How children learn the meanings of words*, Cambridge, MA : MIT Press.
- Bowerman, M., et Choi, S., 2001, Shaping meanings for language : Universal and language-specific in the acquisition of spatial semantic categories, dans M. Bowerman et S. C. Levinson, éd., 2001.
- Bowerman, M., et Levinson, S. C., éd., 2001, *Language acquisition and conceptual development*, Cambridge : Cambridge University Press.
- Boyd, R., 1990, What Realism implies and what it does not, *Dialectica*, 43, 5-29.
- Boyd, R., 1991, Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds, *Philosophical Studies*, 61, 1-2, 127-148.
- Braddon-Mitchell, D., et Jackson, F., 1996, *The philosophy of mind and cognition*, London : Basil Blackwell.
- Brandom, R. B., 2001, *Articulating reasons : An Introduction to inferentialism*, Cambridge, MA : Harvard University Press.
- Brooks, L., 1978, Nonanalytic concept formation and memory for instances, dans E. Rosch et B. B. Lloyd, éd., 1978.
- Bruner, J. S., Goodnow, J.-J., et Austin, G. A., 1956, *A Study of thinking*, New York : Wiley.
- Burge, T., 1979, Individualism and the mental, *Midwest studies in philosophy, vol. 4 : Studies in metaphysics*, 7-121, Minneapolis : University of Minnesota Press.
- Cabeza, R., Bruce. V., Kato, T., et Oda, M., The prototype effect in face recognition : Extension and limits, *Memory and Cognition*, 27, 1, 139-151.

- Cantor, N., et Mischel, W., 1979, Prototypes in person perception, dans L. Berkowitz, éd., *Advances in experimental social psychology*, 12, New York : Academic Press.
- Cantor, N., Mischel, W., et Schwartz, J. C., 1982, Psychiatric diagnosis as prototype categorization, *Journal of Abnormal Psychology*, 14, 45-77.
- Caramazza, A., et Mahlon, B. Z., 2003, The organization of conceptual knowledge : Evidence from category-specific semantic deficits, *Trends in Cognitive Science*, 7, 8, 354-361.
- Carey, S., 1985, *Conceptual change in childhood*, Cambridge, MA : MIT Press.
- Carey, S., 1991, Knowledge acquisition : Enrichment or conceptual change ?, dans S. Carey et R. Gelman, éd., *The Epigenesis of mind: Essays in biology and cognition*, Hillsdale, NJ : Erlbaum.
- Carey, S., et Johnson, S. C., 2000, Metarepresentations and conceptual change : Evidence from Williams syndrome, dans D. Sperber, éd., 2000.
- Carey, S., et Spelke, E. S., 1994, Domain specific knowledge and conceptual change, dans L. Hirschfeld et S. Gelman, éd., 1994.
- Carey, S., et Xu, F., 1999, Sortals and kinds : An appreciation of John Macnamara, dans R. Jackendoff, P. Bloom, et K. Wynn, éd., 1999.
- Carnap, R., 1947/1989, *Meaning and necessity*, Chicago : Chicago University Press.
- Carruthers, P., 1996, *Language, thought and consciousness*, Cambridge : Cambridge University Press.
- Chomsky, N., 1995, Language and nature, *Mind*, 104, 1-61.
- Churchland, P., 1981, Eliminative materialism and the propositional attitudes, *Journal of Philosophy*, 78, 67-90.
- Churchland, P., 1998, Conceptual similarity across sensory and neural diversity : The Fodor/Lepore challenge answered, *Journal of Philosophy*, 95, 1, 5-32.
- Clark, A., 1993, *Associative engines : Connectionism, concepts, and representational change*, Cambridge, MA : MIT Press.
- Clark, A., et Prinz, J. J., 2004, Putting concepts to work : Some thoughts for the twentyfirst century, *Mind and Language*, 19, 1, 57-69.
- Clark, E. V., 1983, Meanings and concepts, dans J. H. Flavell et E. M. Markman, éd., 1983, *Cognitive development*, New York : Wiley.
- Coleman, L., et Kay, P., 1981, Prototype semantics : The english word *lie*, *Language*, 57, 2644.
- Coley, J. D., Medin, D. L., et Atran, S., 1997, Does rank have its privilege ? Inductive inferences within folkbiological taxonomies, *Cognition*, 64, 73-112.

- Coley, J. D., Medin, D. L., Profitt, J. B., Lynch, E., et Atran, Sc., 1999, Inductive reasoning in folkbiological thought, dans D. L. Medin et S. Atran, éd., 1999.
- Collier, J., 1996, On the necessity of natural kinds, dans P. Riggs, éd., 1996, *Natural kinds, laws of nature and scientific reasoning*, Dordrecht : Kluwer.
- Collins, A. M., et Loftus, E. F., 1975, A spreading activation theory of semantic processing, *Psychological Review*, 82, 407-428.
- Conant, M. B., et Trabasso, T., 1964, Conjunctive and disjunctive concept formation under equal-information conditions, *Journal of Experimental Psychology*, 67, 3, 250-255.
- Cosmides, L., 1989, The logic of social exchange : Has natural selection shaped how humans reason ? Studies with the Wason selection task, *Cognition*, 31, 187-276.
- Cosmides, L. et Tooby, J., 1992, Cognitive adaptations for social exchange, dans J. Barkow, L. Cosmides, et J. Tooby, éd., *The Adapted mind: Evolutionary psychology and the generation of culture*. New York : Oxford University Press.
- Costello, F. J., et Keane, M. T., 2000, Efficient creativity : Constraint-guided conceptual combination, *Cognitive Science*, 24, 299–349.
- Costello, F. J., et Keane, M. T., 2001, Testing two theories of conceptual combination : Alignment versus diagnosticity in the comprehension and production of combined concepts, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 255–271.
- Cowie, F., 1999, *What's within, Nativism reconsidered*, New York, NY : Oxford University Press.
- Crane, T, et Mellor, D. H., There is no question of physicalism, *Mind*, 99, 185–206.
- Cummins, R., 1989, *Meaning and mental representation*, Cambridge, MA : MIT Press.
- Cussins, A., 1990, The connectionist construction of concepts, dans M. Boden, éd., 1990, *The Philosophy of artificial intelligence*, Oxford : Oxford University Press.
- Damasio, A. R., 1994, *Descartes' error : Emotion, reason, and the human brain*, New York : Grosset/Putnam.
- Davies, M., 1991, Concepts, connectionism, and the language of thought, dans W. Ramsey, S. Stich, et D. Rumelheart, éd., 1991, *Philosophy and connectionist theory*, Hillsdale, NJ : Erlbaum.
- Davies, S., 2000, Social externalism and deference, dans D. Sperber, éd., 2000.
- Dehaene, S., 1997, *The Number sense, How the mind creates mathematics*, New York : Oxford University Press.
- Dennett, D. C., 1987, *The Intentional stance*, Cambridge, MA : MIT Press.
- Dennett, D. C., 2000, Making tools for thinking, dans D. Sperber, éd., 2000.

- Descartes, R., *Méditations métaphysiques*, 4<sup>ème</sup> éd., Paris : PUF Quadrige.
- Descombes, V., 1995, *La Denrée mentale*, Paris : Editions de Minuit.
- Devitt, M., et Sterelny, K., 1999, *Language and reality, An introduction to the philosophy of language*, 2<sup>e</sup> éd., Cambridge, MA : MIT Press.
- Dupré, J., 1981, Biological taxa as natural kinds, *Philosophical Review*, 90, 66-90.
- Dupré, J., 1986, Sex, gender, and essence, dans P. A. French, Th. E. Uehling et H. K. Wettstein, éd., 1986.
- Dupré, J., 1999, Are whales fish?, dans D. L. Medin et S. Atran, éd., 1999.
- Ereshefsky, M., 1998, Species pluralism and anti-realism, *Philosophy of Science*, 65, 1, 103-120.
- Estes, Z., 2003, Attributive and relational processes in nominal combination, *Journal of Memory and Language*, 48, 304-319
- Evans, G., 1982, *The Varieties of reference*, édité par J. McDowell, Oxford : Clarendon Press.
- Ewald, P. M., 1996, *Evolution of infectious disease*, Oxford : Oxford University Press.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., et Thorpe, S. J., 2000, A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes, *Journal of Cognitive Neuroscience*, 13, 2, 171-180.
- Feyerabend, P., 1975, *Against method*, Londres : Verso.
- Feyerabend, P., 1981, *Problems of empiricism: Philosophical papers, Vol. 2*, Cambridge : Cambridge University Press.
- Field, H., 1973, Theory change and the indeterminacy of reference, *Journal of Philosophy*, 70, 14, 462-481.
- Field, H., 2001, *Truth and the Absence of Fact*, Oxford : Oxford University Press.
- Fisher, S. C., 1916, The Process of generalizing abstraction ; and its product, the general concept, *Psychological Monographs*, XXI, 2, 90, 1-209.
- Fodor, J. A., 1974, Special sciences, *Synthese*, 28, 97-115.
- Fodor, J. A., 1975, *The Language of thought*, New York : Crowell.
- Fodor, J. A., 1981, *Representations : Philosophical essays on the foundations of cognitive science*, Cambridge, MA : MIT Press.
- Fodor, J. A., 1983, *The Modularity of mind*, Cambridge, MA : MIT Press.
- Fodor, J. A., 1987, *Psychosemantics*, Cambridge, MA : MIT Press.
- Fodor, J. A., 1990, *A Theory of content and other essays*, Cambridge, MA : MIT Press.
- Fodor, J. A., 1994, *The Elm and the expert : Mentalese and its semantics*, Cambridge MA : MIT Press.

- Fodor, J. A., 1994, Concepts : A potboiler, *Cognition*, 50, 1, 95-113.
- Fodor, J. A., 1998, *Concepts, Where cognitive science went wrong*, New York : Oxford University Press.
- Fodor, J. A., 1999a, There are no recognitional concepts – not even RED, dans J. A. Fodor 1999c.
- Fodor, J. A., 1999b, There are no recognitional concepts : not even red ; Part 2: The plot thickens, dans J. A. Fodor 1999c.
- Fodor, J. A., 1999c, *In critical condition : Polemical essays on cognitive science and the philosophy of mind*, Cambridge, MA : MIT Press.
- Fodor, J. A., 2000, Replies to critics, *Mind and Language*, 15, 2-3, 350-374.
- Fodor, J. A., 2001, *The Mind doesn't work that way*, Cambridge, MA : MIT Press.
- Fodor, J. A., 2004, Having concepts : A brief refutation of the twentieth century, *Mind and Language*, 19, 1, 29-48.
- Fodor, J. A., Garret, M. F., Walker, E. C. T., et Parkes, C. H., 1980, Against definitions, *Cognition*, 8, 263-367.
- Fodor, J. A., et Lepore, E., 1992, *Holism : A Shopper's guide*, Oxford : Blackwell.
- Fodor, J. A., et Lepore, E., 1996, The red herring and the pet fish : Why concepts still can't be prototypes, *Cognition*, 58, 253-270.
- Fodor, J. A., et McLaughlin, B. P., 1990, Connectionism and the problem of systematicity : Why Smolensky's solution doesn't work, *Cognition*, 35, 183-204.
- Fodor, J. A., et Pylyshyn, Z., 1988, Connectionism and cognitive architecture : A critical analysis, *Cognition*, 28, 3-71.
- Frege, G., 1894/1994, Concept et objet, dans les *Ecrits logiques et philosophiques*, traduit par C. Imbert, 1994, Paris : Seuil.
- French, P. A., Uehling, Th, E, Wettstein, H. K., éd., 1986, *Midwest studies in philosophy, XI, Studies in essentialism*, Minneapolis : University of Minnesota Press.
- Fried, L. S., et Holoyak, K. J., 1984, Induction of category distribution : A framework for classification learning, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 10, 234-257.
- Friedman, M., 1974, Explanation and scientific understanding, *Journal of philosophy*, 71, 5-19.
- Gagné, C. L., et Shoben, E. J., 1997, Influence of thematic relations on the comprehension of modifier-noun combinations, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 23, 71-87.

- Gallistel, C. R., 1999, The replacement of general-purpose learning models with adaptively specialized learning modules, dans M. S. Gazzaniga, éd., *The Cognitive neurosciences*, 2<sup>e</sup> édition, Cambridge, MA : MIT Press.
- Gärdenfors, P., 2000, *Conceptual spaces, The Geometry of thought*, Cambridge, MA : MIT Press.
- Gelman, R., 1990, First principles organize attention to and learning about relevant data : Number and the animate-inanimate distinction as examples, *Cognitive Science*, 14, 79-106.
- Gelman, R., 2001, Cognitive development, dans H. Pashler et D. L. Medin, éd., *Handbook of experimental psychology, Third edition, Vol. 2*, New York : Wiley.
- Gelman, S. A., 1988, The development of induction within natural kinds and artifacts categories, *Cognitive Psychology*, 20, 65-95.
- Gelman, S. A., et Coley, J. D., 1990, The importance of knowing the dodo is a bird : Categories and inference in 2-year-old children, *Developmental Psychology*, 26, 796-804.
- Gelman, S. A., et Coley, J. D., 1991, Language and categorization : The acquisition of natural kind terms, dans S. A. Gelman et J. P. Byrnes, éd., *Perspectives on language and thought*, Cambridge : Cambridge University Press.
- Gelman, S. A., et Hirschfeld, L. A., 1999, How biological is essentialism ?, dans D. L. Medin et S. Atran, éd., 1999.
- Gelman, S. A., et Kremer, K. E., 1991, Understanding natural causes : Children's explanations of how objects and their properties originate, *Child Development*, 62, 396-414.
- Gelman, S. A., et Markman, E., 1986, Categories and induction in young children, *Cognition*, 23, 183-209.
- Gelman, S. A., et Markman, E., 1987, Young children's inductions from natural kinds : The role of categories and appearances, *Child Development*, 58, 1532-1541.
- Gelman, S. A., et Medin, D. L., 1993, What's so essential about essentialism? Different perspective on the interaction of perception, language, and conceptual knowledge, *Cognitive Development*, 8, 157-167.
- Gelman, S. A., et Wellman, H. M., 1991/1999, Insides and essences : Early understandings of the non-obvious, *Cognition*, 38, 213-244, reproduit dans Margolis et Laurence, éd., 1999.
- Gelman, S. A., Coley, J. D., et Gottfried, G. M., 1994, Essentialist beliefs in children : The acquisition of concepts and theories, dans L. A. Hirschfeld et S. A. Gelman, éd., 1994.
- Gigerenzer, G., 1991, How to make cognitive illusions disappear : Beyond « Heuristics and Biases », *European Review of Social Psychology*, 2, 5, 84-115.

- Gigerenzer, G., 1996, On narrow norms and vague heuristics : A reply to Kahneman and Tversky (1996), *Psychological Review*, 103, 3, 592-596.
- Gigerenzer, G., et Hoffrage, U., 1995, How to improve bayesian reasoning without instruction : Frequency formats, *Psychological Review*, 102, 684-704.
- Gigerenzer, G, Todd, P. M., et the ABC Research Group, 1999, *Simple heuristics that make us smart*, Oxford : Oxford University Press.
- Gil-White, F., 1999, How thick is blood? The plot thickens ... : If ethnic actors are primordialists, what remains of the circumstantialists/primordialists controversy?, *Ethnic and Racial Studies*, 22, 5, 789-820.
- Gil-White, F., 2001, Are ethnic group biological « species » to the human brain, *Current Anthropology*, 42, 4, 515-455.
- Giroto, V., Kimmelmeier, M., Sperber, D., et van der Henst, J.-B., 2001, Inept reasoners or pragmatic virtuosos ? Relevance and the deontic selection task, *Cognition*, 81, 2, B69-B76.
- Goldstein, D. G., et Gigerenzer, G., 2002, Models of ecological rationality : The recognition heuristic, *Psychological Review*, 109, 1, 75-90.
- Goldstone, R. L., et Barsalou, L. W., 1998, Reuniting perception and conception, *Cognition*, 65, 231-262.
- Goodman, N., 1955, *Fact, fiction, and forecast*, Indianapolis : Bobbs-Merrill.
- Gopnik, A., et Meltzoff, A. N., 1997, *Words, thoughts, and theories*, Cambridge, MA : MIT Press.
- Gopnik, A., et Wellman, H. M., 1994, The theory theory, dans L. A. Hirshfeld et S. Gelman, éd., 1994.
- Goschke, T., 1997, Implicit learning and unconscious knowledge : Mental representations, computational mechanisms and brain structure, dans K. Lamberts et D. Shanks, éd., 1997.
- Greene, J. D., et Haidt, J., 2002, How (and where) does moral judgment work ?, *Trends in Cognitive Science*, 6, 517-523.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., et Cohen, J. D., 2001, An fMRI investigation of emotional engagement in moral judgment, *Science* 14, 293, 2105-1208.
- Grice, P., 1989, *Studies in the way of words*, Cambridge, MA : Harvard University Press.
- Griffith, P. E., 1997, *What emotions really are*, Chicago : Chicago University Press.
- Griffiths, P. E., 2003, Emotions as natural kinds and normative kinds, disponible sur : <http://philsci-archive.pitt.edu/>.
- Grünbaum, A., 1908, Über die Abstraktion der Gleichheit, *Archive für die Geschichte der Psychologie*, 340-478.

- Hacking, I., 1991, A tradition of natural kinds, *Philosophical Studies*, 61, 109-126.
- Hahn, U., et Chater, N., 1997, Concepts and similarity, dans K. Lamberts et D. Shanks, éd., 1997.
- Hahn, U., et Chater, N., 1998, Similarity and rules : distinct ? exhaustive ? empirically distinguishable ?, *Cognition*, 65, 197-230.
- Hahn, U., et Ramscar, M., éd., 2001, *Similarity and categorization*, Oxford : Oxford University Press.
- Haidt, J., 2001, The emotional dog and its rational tail : A social intuitionist approach to moral judgment, *Psychological Review*, 108, 814-834.
- Hammond, K., 1989, *Case-based planning : Viewing planning as a memory task*, New York : Academic Press.
- Hampton, J. A., 1979, Polymorphous concepts in semantic memory, *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.
- Hampton, J. A., 1981, An Investigation of the nature of abstract concepts, *Memory and Cognition*, 9, 149-156.
- Hampton, J. A., 1982, A demonstration of intransitivity in natural categories, *Cognition*, 12, 151-164.
- Hampton, J. A., 1987, Inheritance of attributes in natural concept conjunctions, *Memory and Cognition*, 15, 55-71.
- Hampton, J. A., 1988, Overextension of conjunctive concepts : Evidence for a unitary model of concept typicality and class inclusion, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 14, 378-383.
- Hampton, J. A., 1993, Prototype models of concept representation, dans I. Van Mechelen, J. A., Hampton, R. S. Michalski, et P. Theuns, éd., *Categories and concepts : Theoretical views and inductive data analysis*, Londres : Academic Press.
- Hampton, J. A., 1995, Testing the prototypes theory of concepts, *Journal of Memory and Language*, 34, 686-708.
- Hampton, J. A., 1996, Conjunctions of visually-based categories : Overextensions and compensation, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 34, 686-708.
- Hampton, J. A., 1997a, Psychological representation of concepts, dans M. A. Conway, éd., *Cognitive Models of Memory*, Cambridge, MA: MIT Press.
- Hampton, J. A., 1997b, Conceptual combination, dans K. Lamberts et D. Shanks, éd., 1997.

- Hampton, J. A., 1997c, Conceptual combination : Conjunction and negation of natural concepts, *Memory and Cognition*, 25, 888-909.
- Hampton, J. A., 1998, Similarity-based categorization and fuzzyness of natural categories, *Cognition*, 65, 137-165.
- Hampton, J. A., 2000, Concepts and prototypes, *Mind and Language*, 15, 2-3, 299-307.
- Hampton, J. A., 2001, The role of similarity in natural categorization, dans U. Hahn et M. Ramscar, éd., 2001.
- Hatano, G., et Inagaki, K., 1999, A developmental perspective on informal biology, dans D. L. Medin et S. Atran, éd., 1999.
- Hauser, M. D., Newport, E. L., and Aslin, R. N., 2001, Segmentation of the speech stream in a nonhuman primate : Statistical learning in cotton-top tamarins, *Cognition*, 78, B53-B64.
- Hegel, G. W., 1997, *La Phénoménologie de l'esprit*, traduit par J.-P. Lefebvre, Paris : Aubier.
- Heit, E., 1997, Knowledge and concept learning, dans K. Lamberts et D. Shanks, éd., 1997.
- Heit, E., 2001, Background Knowledge and models of categorization, dans U. Hahn et M. Ramscar, éd., 2001.
- Hertwig, R., et Gigerenzer, G., 1999, The « conjunction fallacy » revisited : How intelligent inferences look like reasoning errors, *Journal of Behavioral Decision Making*, 12, 275-305.
- Hintzman, D. L., 1986, Schema abstraction in a multiple-trace memory model, *Psychological Review*, 93, 411-428.
- Hintzman, D. L., et Ludlam, G., 1980, Differential forgetting of prototypes and memory for instances : Memory, perception, and concept learning, *Memory and Cognition*, 8, 378-382.
- Hirschfeld, L. A., 1996, *Race in the making : Cognition, culture, and the child's construction of human kinds*, Cambridge, MA : MIT Press.
- Hirschfeld, L. A., et Gelman, S. A., 1994, *Mapping the mind : Domain-specificity in cognition and culture*, New York : Cambridge University Press.
- Hodges, W., 1998, Compositionality is not the problem, *Logic and Philosophy*, 6, 7-33.
- Hodges, W., 2001, Formal features of compositionality, *Journal of Logic, Language and Information*, 10, 1, 7-28.
- Holyoak, K. J., et Glass, A. L., 1975, The role of contradictions and counterexamples in the rejection of false sentences, *Journal of Verbal Learning and Verbal Behavior*, 14 : 215-239.
- Homa, D., Sterling, S., et Trepel, L., 1981, Limitations of exemplar-based generalization and the abstraction of categorical information, *Journal of Experimental Psychology : Human Learning and Memory*, 7, 6, 418-439.

- Hull, C. L., 1920, Quantitative aspects of the evolution of concepts, *Psychological Monographs*, XXVIII.
- Hull, D., 1965, The effects of essentialism on taxonomy : Two thousand years of stasis, *The British Journal for the Philosophy of Science*, 15, 314-326 et 16,1-18.
- Hull, D., 1978, A matter of individuality, *Philosophy of Science*, 45, 335-360.
- Hume, D., 1739/1978, *A Treatise of human nature*, édité par L. A. Selby-Bigge et P. H. Nidditch, Oxford : Oxford University Press.
- Hume, D., 1748/1975, *Enquiries concerning human understanding and concerning the principles of morals*, édité par P. H. Nidditch, Oxford : Oxford University Press.
- Huttenlocher, J., et Hedges, L. V., 1994, Combining graded categories : Membership and typicality, *Psychological Review*, 101, 1, 157-165.
- Inhelder, B., et Piaget, J., 1969, *The Early growth of logic in the child : Classification and seriation*, New York : W. W. Norton.
- Jackendoff, R., 1992a, What is a concept, that a person may grasp it ?, dans R. Jackendoff 1992b.
- Jackendoff, R., 1992b, *Languages of the mind : Essays on mental representation*, Cambridge, MA : MIT Press.
- Jackendoff, R., 1995, *Patterns in the mind : Language and human nature*, New York : Basic Books.
- Jackendoff, R., Bloom, P., et Wynn, K., éd., 1999, *Language, logic, and concepts*, Cambridge, MA : MIT Press.
- Jackson, F., 1997, *From metaphysics to ethics*, Oxford : Oxford University Press.
- Jackson, F., 1998, Reference and description revisited, *Philosophical Perspectives*, 12, *Language, Mind, and Ontology*, 201-218.
- Jackson, F., et Pettit, P., 1988, Functionalism and Broad Content, *Mind*, 97, 381-400.
- Jacob, P., 1997, *Pourquoi les choses ont-elles un sens ?*, Paris : Odile Jacob.
- Jacob, P., et Jeannerod, M., 2003, *Ways of seeing : The Scope and limits of visual cognition*, Oxford : Oxford University Press.
- Janssen, M. V. Th., 1997, Compositionality, dans J. van Benthem et A. ter Meulen, éd., *Handbook of logic and language*, Elsevier Science B.V.
- Jeyifous, S., 1986, *Atimodem : Semantic conceptual development among the Yoruba*, Doctoral dissertation, Cornell University.
- Johnson, C. et Keil, F. C., 2000, Explanatory understanding and conceptual combination, dans F. C. Keil et R. A. Wilson, éd., 2000.

- Jones, S. S., et Smith, L. B., 1993, The place of perception in children's concepts, *Cognitive Development*, 8, 113-139.
- Kahneman, D., et Tversky, A., 1972, Subjective probability : A judgment of representativeness, *Cognitive Psychology*, 3, 430-454.
- Kamp, H., et Partee, B., 1995, Prototype theory and compositionality, *Cognition*, 57, 129-191.
- Kant, E., 2001, *La Critique de la raison pure*, traduction de A. Tremesaygues et B. Pacaud, sixième éd., Paris : PUF.
- Katz, J. J., et Fodor, J. A., 1963, The structure of a semantic theory, *Language*, 39, 170-210.
- Keil, F. C., 1987, Category structure and patterns of conceptual change, dans U. Neisser, éd., 1987.
- Keil, F. C., 1989, *Concepts, kinds, and cognitive development*, Cambridge, MA : MIT Press.
- Keil, F. C., 1991, Theories, concepts, and word meaning, dans S. A. Gelman et J. P. Byrnes, éd., *Perspective on language and thought, interrelations in development*, Cambridge : Cambridge University Press.
- Keil, F. C., 1994, The birth and nurturance of concepts by domains : The origins of concepts of living things, dans L. A. Hirschfeld et S. A. Gelman, éd., 1994.
- Keil, F. C., et Batterman, N., 1984, A characteristic-to-defining shift in the acquisition of word meaning, *Journal of Verbal Learning and Verbal Behavior*, 23, 221-236.
- Keil, F. C., Carter Smith, W., Simons, D. J., et Levin, D. T., 1998, Two dogmas of conceptual empiricism : implications for hybrid models of the structure of knowledge, *Cognition*, 65, 103-135.
- Keil, F. C., Levin, D. T., Richman, B. A., et Gutheil, G., 1999, Mechanism and explanation in the development of biological thought : The case of disease, dans D. L. Medin et S. Atran, éd., 1999.
- Keil, F. C., et Richardson, D. C., 1999, Species, stuff, et patterns of causation, dans Robert A. Wilson, éd., *Species: New interdisciplinary papers*, Cambridge, MA : MIT Press.
- Keil, F. C., et Wilson, R. A., éd., 2000, *Explanation and cognition*, Cambridge, MA : MIT Press.
- Kelly, R., L., 1995, *The Foraging spectrum*, Washington : Smithsonian Institution Press.
- Kim, J., 1993, *Supervenience and mind: Selected philosophical essays*, Cambridge : Cambridge University Press.
- Kim, J., 1998, *Mind in a physical world: An essay on the mind-body problem and mental causation*, Cambridge, MA : MIT Press.

- Kitchener, E. G. et Squire, L. R., 2000, Impaired verbal category learning in amnesia, *Behavioral Neuroscience*, 114, 5, 907-911.
- Kitcher, P., 1981, Explanatory unification, *Philosophy of Science*, 48, 507-531.
- Kitcher, P., 1993, *The Advancement of science*, Oxford : Oxford University Press.
- Knowlton, B. J., 1997, Declarative and nondeclarative knowledge : Insights from cognitive neuroscience, dans K. Lamberts et D. R. Shanks, éd., 1997.
- Knowlton, B. J., 1999, What can neuropsychology tell us about category learning, *Trends in Cognitive Science*, 3, 4, 123-124.
- Knowlton, B. J., et Squire, L. R., 1993, The learning of categories : Parallel brain systems for item memory and category knowledge, *Science*, 262, 1747-1749.
- Kolodny, J. A., 1994, Memory processes in classification learning : An investigation of amnesic performance in categorization of dot patterns and artistic styles, *Journal of Verbal Learning and Verbal Behavior*, 23, 734-759.
- Komatsu, L. K., 1992, Recent views of conceptual structure, *Psychological Bulletin*, 112, 500-526.
- Kosslyn, S. M., 1978, Imagery and internal representations, dans E. Rosch et B. Lloyd, éd., 1978.
- Kripke, S. A., 1972/1982, *Naming and necessity*, Cambridge, MA : MIT Press.
- Kruschke, J. K., ALCOVE : An exemplar-based connectionist model of category learning, *Psychological Review*, 99, 22-44.
- Kuhn, T. S., 1970, *The Structure of scientific revolutions*, 2<sup>e</sup> éd., Chicago, IL : University of Chicago Press.
- Kunda, Z., Miller, D. T., et Claire, T., 1990, Combining social concepts : The role of causal reasoning, *Cognitive Science*, 14, 551-577.
- Lamberts, K., 2000, Information-accumulation theory of speeded categorization, *Psychological Review*, 107, 227-260.
- Lamberts, K., et Shanks, D., éd., 1997, *Knowledge, concepts and categories*, Cambridge, MA : MIT Press.
- Laudan, L., 1981, A confutation of convergent realism, *Philosophy of Science*, 48, 19-48.
- Laurence, S., et Margolis, E., 1999, Concepts and cognitive science, dans E. Margolis et S. Laurence, éd., 1999.
- Laporte, J., 1998, Living water, *Mind*, 107, 451-455.
- Lepore, E., et Fodor, J. A., 1999, All at sea in semantic space : Churchland on meaning similarity, *The Journal of Philosophy*, 96, 381-403.

- Leslie, A. M., 2000, How to acquire a representational theory of mind, dans D. Sperber, éd., 2000.
- Leslie, A. M., Xu, F., Tremoulet, P., et Scholl, B., 1998, Indexing and the object concept: Developing 'what' and 'where' systems, *Trends in Cognitive Sciences*, 2, 10–18.
- Levinson, S. C., 2001, Covariation between spatial language and cognition, and its implication for language learning, dans M. Bowerman et S. C. Levinson, éd., 2001.
- Lewis, D. K., 1979, How to Define Theoretical Terms, *Journal of Philosophy*, 67, 427-446.
- Lewis, D. K., 1983, New work for a theory of universals, *Australian Journal of Philosophy*, 61, 343-77.
- Libera, de, A., 1997, *La Querelle des universaux*, Paris : Seuil.
- Locke, J., 1690/1964, *An Essay concerning human understanding*, A. D. Woozley, éd., New York : Meridian.
- Lyons, J. C., 2001, Carving the mind at its (not necessarily modular) joints, *The British Journal for the Philosophy of Science*, 52, 277-302.
- Machery, E., 2002, Compositionnalité, composition conceptuelle, combinaison prototypique, *In Cognito, Revue Francophone Internationale en Sciences Cognitives*, 22, 5-26.
- Machery, E., 2004, Catégorisation et attribution de propriétés, *Les Cahiers Philosophiques de Strasbourg*, sous presse.
- Machery, E., Mallon, R., Nichols, S., et Stich, S. P., 2004, Semantics, Cross-cultural style, *Cognition*, 92, 3, B1-B12.
- Mahon, B. Z., et Caramazza, A., 2003, There are facts... And then there are facts, A reply to Moss and Tyler, *Trends in Cognitive Science*, 7, 11, 481-482.
- Malt, B. C., 1989, An on-line investigation of prototype and exemplar strategies in classification, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 15, 4, 539-555.
- Malt, B. C., 1993, Concept structure and category boundaries, dans G. Nakamura, R. Taraban, et D. Medin, éd., 1993, *The Psychology of learning and motivation, vol. 29, Categorization by humans and machines*, Orlando, FL : Academic Press, Inc.
- Malt, B. C., 1994, Water is not H<sub>2</sub>O, *Cognitive Psychology*, 27, 41-70.
- Malt, B. C., et Johnson, E. C., 1992, Do artifact concepts have cores ?, *Journal of Memory and Language*, 31, 195-217.
- Malt, B. C., et Smith, E. E., 1984, Correlated properties in natural categories, *Journal of Verbal Learning and Verbal Behavior*, 23, 250-269.

- Mandler, J. M., 1992, How to build a baby : II. Conceptual primitives, *Psychological Review*, 99, 587-604.
- Mandler, J. M., Bauer, P. J., et McDonough, L., 1991, Separating the sheep from the goats : Differentiating global categories, *Cognitive Psychology*, 23, 263-298.
- Mandler, J. M., et McDonough, L., 1996, Drinking and driving don't mix : Inductive generalizations in infancy, *Cognition*, 59, 307-318.
- Marcus, G. F., *The Algebraic mind : Integrating connectionism and cognitive science*, Cambridge, MA : MIT Press.
- Margolis, E., 1994, A reassessment of the shift from the classical theory of concepts to prototype theory, *Cognition*, 51, 1, 73-89.
- Margolis, E., 1995, The significance of the theory analogy in the psychological study of concepts, *Mind and Language*, 10, 45-71.
- Margolis, E., et Laurence, S., éd., 1999, *Concepts, Core readings*, Cambridge, MA : MIT Press.
- Markman, E., 1989, *Categorization and naming in children : Problems in induction*, Cambridge MA : MIT Press.
- Marr, D., 1982, *Vision : A computational investigation into the human representation and processing of visual information*, San Francisco, CA : W. H. Freeman.
- McCloskey, M., et Glucksberg, S., 1978, Natural categories : Well-defined or fuzzy sets ?, *Memory and Cognition*, 6, 462-472.
- McDowell, J., 1994, *Mind and world*, Cambridge, MA : Harvard University Press.
- McNamara, T. P., et Sternberg, R. J., 1983, Mental models of word meaning, *Journal of Verbal Learning and Verbal Behavior*, 22, 449-474.
- Medin, D. L., 1989, Concepts and conceptual structure, *American Psychologist*, 44, 1469-1481.
- Medin, D. L., et Atran, S., éd., 1999, *Folkbiology*, Cambridge, MA : MIT Press.
- Medin, D. L., et Edelson, S. M., 1988, Problem structure and the use of base-rate information from experience, *Journal of Experimental Psychology : General*, 117, 1, 68-85.
- Medin, D. L., Goldstone, R. L., et Gentner, D., 1993, Respects for similarity, *Psychological Review*, 100, 254-278.
- Medin, D. L., et Ortony, A., 1989, Psychological essentialism, dans S. Vosniadou et A. Ortony, éd., 1989.
- Medin, D. L., et Schaffer, M. M., 1978, Context theory of classification learning, *Psychological Review*, 85, 207-238.

- Medin, D. L., et Shoben, E. J., 1988, Context and structure in conceptual combination, *Cognitive Psychology*, 20, 158-190.
- Medin, D. L., et Schwanenflugel, P. J., 1981, Linear separability in classification learning, *Journal of Experimental Psychology : Human Learning and Memory*, 7, 355-368.
- Medin, D. L., et Wattenmaker, W. D., 1987, Category cohesiveness, theories, and cognitive archeology, dans U. Neisser, éd., 1987.
- Medin, D. L., Goldstone, R. L., et Gentner, D., 1993, Respects for similarity, *Psychological Review*, 100, 2, 254-278.
- Mellor, J. S., 1977, Natural kinds, *The British Journal for the Philosophy of Science*, 28, 299-312.
- Mill, J. S., 1905, *A System of logic*, Londres : Routledge.
- Miller, G. A., et Johnson-Laird, P. N., 1976, *Language and perception*, Cambridge, MA : Harvard University Press.
- Millikan, R. G., 1984, *Language, thought, and other biological categories : New foundations for realism*, Cambridge MA : MIT Press.
- Millikan, R. G., 1993, *White queen psychology and other essays for Alice*, Cambridge, MA : MIT Press.
- Millikan, R. G., 1998, A common structure for concepts of individuals, stuffs, and real kinds : More mama, more milk, and more mouse, *Behavioral and Brain Sciences*, 21, 55-100.
- Millikan, R. G., 2000, *On clear and confused ideas : An essay about substance concepts*, New York : Cambridge University Press.
- Mitchell, S. D., 2003, *Biological complexity and integrative pluralism*, Cambridge : Cambridge University Press.
- Moore, T., 1910, The process of abstraction : An experimental study, *University of California Publications in Psychology*, I, 2.
- Moss, H. E., et Tyler, L. K., 2003, Weighing up the facts of category-specific semantic deficits, *Trends in Cognitive Science*, 7, 11, 480-481.
- Murphy, G. L., 1988, Comprehending complex concepts, *Cognitive science*, 12, 529-562.
- Murphy, G. L., 1990, Noun phrase interpretation and conceptual combination, *Journal of Memory and Language*, 29, 259-288.
- Murphy, G. L., 1993, A rational theory of concepts, dans G. V. Nakamura, R. M. Taraban, et D. L., Medin, éd., *The Psychology of learning and motivation, Vol. 29 : categorization by humans and machine*, New York : Academic Press.
- Murphy, G. L., 2002, *The Big book of concepts*, Cambridge, MA : MIT Press.

- Murphy, G. L., et Allopena, P. D., 1994, The locus of knowledge effects in concept learning, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 20, 4, 904-919.
- Murphy, G. L., et Lassaline, M. E., 1997, Hierarchical structure in concepts and the basic level of categorization, dans K. Lamberts et D. Shanks, 1997, éd.
- Murphy, G. L., et Medin, D. L., 1985/2004, The role of theories in conceptual coherence, *Psychological Review*, 92, 289-316, traduit par E. Machery et F. Athané, *Cahiers de Philosophie de Strasbourg*, sous presse.
- Neisser, U., éd., 1987, *Concepts and conceptual development : Ecological and intellectual factors in categorization*, Cambridge : Cambridge University Press.
- Nichols, S., Stich, S. P., et Weinberg, J. M., 2003, Meta-skepticism : Meditations in ethno-epistemology, dans S. Luper, éd., *The Sceptics*, Aldershot : Ashgate Publishing.
- Nisbett, R. E., 2003, *The Geography of thought, How Asians and Westerners think Differently... and why*, New York, NY : The Free Press.
- Nisbett, R. E., Peng, K., Choi, I., et Norenzayan, A., 2001, Culture and systems of thought : Holistic vs. analytic cognition, *Psychological Review*, 108, 291-310.
- Nisbett, R. E., et Wilson, T., 1977, Telling more than we can know : Verbal reports on mental processes, *Psychological Review*, 8, 231-259.
- Noe, A., et O'Reagan, K., 2001, A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences* 24, 5, 883-975
- Norenzayan, A., Smith, E. E., Kim, B., J., et Nisbett, R. E., 2003, Cultural preferences for formal versus intuitive reasoning, *Cognitive Science*, 26, 653-684.
- Norton, J. D., 2001, A material theory of induction, *Philosophy of Science*, 70, 647-670.
- Nosofsky, R. M., 1986, Attention, similarity, and the identification-categorization relationship, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 10, 104-114.
- Nosofsky, R. M., 1988, Exemplar-based accounts of relations between classification, recognition, and typicality, *Journal of Experimental Psychology : Learning, Memory and Cognition*, 14, 700-708.
- Nosofsky, R. M., 1992, Exemplar-based approach to relating categorization, identification, and recognition, dans F. G. Ashby, éd., *Multidimensional models of perception and cognition*, Hillsdale, NJ : Lawrence Erlbaum Associates.
- Nosofsky, R. M., 2000, Exemplar representation without generalization? Comment on Smith and Minda's (2000) « Thirty categorization results in search of a model », *Journal of Experimental Psychology : Learning, Memory and Cognition*, 26, 1735-1743.

- Nosofsky, R. M., et Zaki, S. R., 1998, Dissociations between categorization and recognition in amnesic and normal individuals : An exemplar-based interpretation, *Psychological Science*, 9, 247-255.
- Nosofsky, R. M., Palmeri, T. J., et McKinley, S. C., 1994, Rule-plus-exception model of classification learning, *Psychological Review*, 101, 266-300.
- Okasha, S., 2003, Does the concept of 'clade selection' make sense ?, *Philosophy of Science*, 70, 739–751.
- Osherson, D. N., et Smith, E. E., 1981, On the adequacy of prototype theory as a theory of concepts, *Cognition*, 9, 35-58.
- Osherson, D. N., et Smith, E. E., 1997, On typicality and vagueness, *Cognition*, 64, 189-206.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., Shafir, E., 1990, Category-based induction, *Psychological review*, 97, 185-200.
- Pacherie, E., 1993, *Naturaliser l'intentionnalité, Essai de philosophie de la psychologie*. Paris : PUF.
- Pacherie, E., 1997, Faut-il avoir peur du holisme sémantique, dans D. Dubois, éd., *Catégorisation, représentation et systèmes symboliques*, Lyon : Kimé.
- Panaccio, C., 2002, *Les mots, les concepts et les choses*, Paris : Vrin.
- Peacocke, C., 1986, Explanation in computational psychology: Language, perception, and level 1.5, *Mind and Language*, 1, 2, 173-193.
- Peacocke, C., 1992, *A Study of concepts*, Cambridge, MA : MIT Press.
- Peacocke, C., 1996, Can possession conditions individuate concepts?, *Philosophy and Phenomenological Research*, LVI, 433-460.
- Peacocke, C., 1998a, Implicit conceptions, understanding and rationality, dans E. Villanueva, éd., 1998.
- Peacocke, C., 1998b, Implicit conceptions, the a priori, and the identity of concepts, dans E. Villanueva, éd., 1998.
- Peacocke, C., 1999, *Being known*, Oxford : Oxford University Press.
- Peacocke, C., 2000, Fodor on concepts : Philosophical aspects, *Mind and Language*, 15, 2-3, 327-340.
- Peacocke, C., 2004, Interrelations : Concepts, knowledge, reference and structure, *Mind and Language*, 19, 1, 85-98.
- Perner, J., 1991, *Understanding the representational mind*, Cambridge, MA : MIT Press.
- Piaget, J., 1953, *La Psychologie de l'intelligence*, Paris : A. Colin.

- Pinker, S., et Prince, A., 1988, On language and connectionism : Analysis of a parallel distributed processing model of language acquisition, *Cognition*, 28, 73-193.
- Pinker, S., et Prince, A., 1999, The nature of human concepts: Evidence from an unusual source, dans R. Jackendoff, P. Bloom, et K. Wynn, éd., 1999.
- Platon, 1993, *La République*, traduit par P. Tachet, Paris : Folio Essais.
- Posner, M. I., et Keele, S. W., 1968, On the genesis of abstract ideas, *Journal of Experimental Psychology*, 77, 353-363.
- Posner, M. I., et Keele, S. W., 1970, Retention of abstract ideas, *Journal of Experimental Psychology*, 83, 304-308.
- Prasada, S., 1999, Names for things and stuff: An aristotelian perspective, dans R. Jackendoff, P. Bloom, et K. Wynn, éd., 1999.
- Prinz, J. J., 2002, *Furnishing the mind: Concepts and their perceptual basis*, Cambridge, MA : MIT Press.
- Putnam, H., 1967/1975, The nature of mental states, dans H. Putnam, *Mind, language, and reality*, vol. 2, Londres : Cambridge University Press.
- Putnam, H., 1975, The meaning of « meaning », dans H. Putnam, *Mind, language, and reality*, vol. 2, Londres : Cambridge University Press.
- Quine, W. V. O., 1951, Two dogmas of empiricism, *Philosophical Review*, 60, 20-43.
- Quine, W. V. O., 1960, *Word and Object*, Cambridge, MA : MIT Press.
- Quine, W. V. O., 1977, Natural kinds, dans S. P. Schwartz, éd., 1977, *Naming, necessity, and natural kinds*, Ithaca, NY : Cornell University Press.
- Ramsey, W., Stich, S. P., et Garon, J., 1989, Connectionism, eliminativism and the future of folk psychology, *Philosophical Perspectives*, 4 : *Action Theory and Philosophy of Mind*, 499-533.
- Reber, P. J., Stark, C. E. L., et Squire, L. R., 1998, Cortical areas supporting category learning identified using functional MRI, *Proceedings of the National Academy of Sciences of the United States of America*, 95, 747-750.
- Récanati, F., 2002, The Fodorian fallacy, *Analysis*, 62, 4, 285-289.
- Rey, G., 1983, Concepts and stereotypes, *Cognition*, 15, 237-262.
- Rey, G., 1985, Concepts and conceptions: A reply to Smith, Medin and Rips, *Cognition*, 19, 297-303.
- Rey, G., éd., 1991, *Meaning in mind : Fodor and his critics*, Londres : Blackwell.
- Rips, L. J., 1989, Similarity, typicality, and categorization, dans S. Vosniadou et A. Ortony, éd. 1989.

- Rips, L. J., 1995, The current status of the research on concept combination, *Mind and Language*, 10, 72-104.
- Rips, L. J., Shoben, E. J., et Smith, E. E., 1973, Semantic distance and the verification of semantic relations, *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Rosch, E., 1973, On the internal structure of perceptual and semantic categories, dans T. Moore, éd., *Cognitive development and acquisition of language*, New York : Academic Press.
- Rosch, E., 1975, Cognitive representations of semantic categories, *Journal of Experimental Psychology : General*, 104, 192-233.
- Rosch, E., 1978, Principles of categorization, dans E. Rosch et B. B. Lloyd, éd., 1978.
- Rosch, E., et Lloyd, éd., 1978, *Cognition and categorization*, Hillsdale, NJ : Erlbaum.
- Rosch, E., et Mervis, C. B., 1975, Family resemblance : Studies in the internal structure of categories, *Cognitive Psychology*, 7, 573-605.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D., et Boyes-Braem, P., 1976, Basic objects in natural categories, *Cognitive Psychology*, 7, 573-605.
- Ross, B. H., Perkins, S. J., et Tenpenny, P. L., 1990, Reminding-based category learning, *Cognitive Psychology*, 22, 460-492.
- Ross, B. H., et Makin, V. S., 1999, Prototype versus exemplar models in cognition, dans R. J. Sternberg, éd., *The Nature of cognition*, Cambridge MA : MIT Press.
- Rumelhart, D. E., 1980, Schemata : The building blocks of cognition, dans R. J. Spiro, B. C. Bruce, et W. F. Brewer, éd., *Theoretical issues in reading comprehension*, Hillsdale, NJ : Erlbaum.
- Russell, B., 1948, *Human knowledge :: Its scope and its limits*, Londres : Routledge.
- Saffran, J. R., Aslin, R. N., et Newport, E. L., 1996, Statistical learning by 8-month-old infants, *Science*, 274, 1926-1928.
- Schank, R. C., et Abelson, R. P., 1977, *Scripts, plans, goals and understanding*, Hillsdale, NJ : Erlbaum.
- Schwartz, B., et Reisberg, D., 1991, *Learning and memory*, New York : Norton.
- Schwartz, S. P., 1978, Putnam on artifacts, *Philosophical Review*, 87, 566-574.
- Schwartz, S. P., 1980, Natural kind terms, *Cognition*, 7, 301-315.
- Searle, J. R., 1958, Proper names, *Mind*, 67, 166-173.
- Shafir, E. B., Smith, E. E., et Osherson, D. N., 1990, Typicality and reasoning fallacies, *Memory and Cognition*, 18, 3, 229-239.
- Shank, R. C., Collins, G. C., et Hunter, L. E., 1986, Transcending inductive category formation in learning, *Behavioral and Brain Sciences*, 9, 639-686.

- Shanks, D. R., 1997a, Representation of categories and concepts in memory, dans M. A. Conway, éd., *Cognitive Models of Memory*, Cambridge, MA : MIT Press.
- Shanks, D. R., 1997b, Distributed representations and implicit knowledge : A brief introduction, in K. Lamberts et D. Shanks, éd., 1997.
- Shaver, P. R., Shelley, W, et Schwartz, J. C., 1992, Cross-cultural similarities and differences in emotion and its representation : A prototype approach, dans M. S., Clark, éd., *Emotion, Vol. 13 of Review of Personality et Psychology*, London : Sage.
- Simons, D. J., et Keil, F. C., 1995, An abstract to concrete shift in the development of biological thought : The insides story, *Cognition*, 56, 129-163.
- Sloman, S. A., 1993, Feature-based induction, *Cognitive Psychology*, 25, 231-280.
- Sloman, S. A., 1996, The empirical case for two systems of reasoning, *Psychological Bulletin*, 119, 3-22.
- Sloman, S. A., 1998, Categorical inference is not a tree : The myth of inheritance hierarchies, *Cognitive Psychology*, 35, 1-33.
- Sloman, S. A., et Rips, L. J., 1998, Similarity as an explanatory construct, *Cognition*, 65, 87-101.
- Sloman, S. A., Malt, B. C., et Fridman, A., 2001, Categorization versus similarity : The case of container names, dans U. Hahn et M. Ramscar, éd., 2001.
- Smith, E. E., 1989, Concepts and induction, dans M. Posner, éd., *Foundations of cognitive science*, Cambridge, MA : MIT Press.
- Smith, E. E., 1995, Concepts and reasoning, dans E. E. Smith et D. N. Osherson, éd., 1995, *An Invitation to cognitive science, Vol. 3, 2<sup>nd</sup> ed.*, Cambridge, MA : MIT Press.
- Smith, E. E., et Jonides, J., 1999, The cognitive neuroscience of categorization, dans M. Gazzaniga, éd., *The new cognitive neurosciences, 2<sup>nd</sup> edition*, Cambridge, MA : MIT Press.
- Smith, E. E., et Medin, D. L., 1981, *Categories and concepts*, Cambridge, MA : Harvard University Press.
- Smith E. E., et Osherson, D. N., 1989, Similarity and decision making, dans S. Vosniadou et A. Ortony, éd. 1989.
- Smith, E. E., et Sloman, S. A., 1994, Similarity- versus rule-based categorization, *Memory and Cognition*, 22, 4, 377-386.
- Smith, E. E., Langston, C., et Nisbett, R. E., 1992, The case for rule in reasoning, *Cognitive Science*, 16, 1-40.
- Smith, E. E., Medin, D. L., et Rips, L. J., 1984, A psychological approach to concepts : Comments on Georges Rey's « Concepts and stereotypes », *Cognition*, 17, 265-274.

- Smith, E. E., Osherson, D. N., Rips, L. J., et Keane, M., 1988, Combining concepts : A selective modification model, *Cognitive Science*, 12, 485-527.
- Smith, E. E., Patalano, A. L., et Jonides, J., 1998, Alternative strategies of categorization, *Cognition*, 65, 167-196.
- Smith, E. E., Shoben, E. J., et Rips, L. J., 1974, Structure and process in semantic memory : A featural model for semantic decisions, *Psychological Review*, 81, 214-241.
- Smith, J. D., et Minda, J. P., 1998, Prototypes in the mist : The early epochs of category learning, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 24, 1411-1436.
- Smith, J. D., et Minda, J. P., 2000a, Thirty categorization results in search of a model, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 26, 3-27.
- Smith, J. D., et Minda, J. P., 2000b, Journey to the center of the category : The dissociation in amnesia between categorization and recognition, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 27, 984-1002.
- Smoke, K. L., 1932, An objective study of concept formation, *Psychological Monographs*, XLII.
- Smolensky, P., 1988, On the proper treatment of connectionism, *Behavioral and Brain Sciences*, 11, 1-23.
- Smolensky, P., 1990, Tensor product variable binding and the representation of symbolic structures in connectionist systems, *Artificial Intelligence*, 46, 159-216.
- Smolensky, P., 1991, Connectionism, constituency, and the language of thought, dans G. Rey, éd., 1991.
- Sober, E. R., 1994, Let's razor Ockham's razor, dans E. Sober, *From a biological point of view : Essays in evolutionary philosophy*, Cambridge : Cambridge University Press.
- Sober, E. R., et Wilson, D. S., 1998, *Unto others : The evolution and psychology of unselfish behavior*, Cambridge, MA : Harvard University Press.
- Sober, E. R., 1999, *Philosophy of biology*, 2<sup>e</sup> éd., Oxford : Oxford University.
- Spalding, T. L., et Ross, B. H., 1994, Comparison-based learning : Effects of comparing instances during category learning, *Journal of Experimental psychology : Learning, Memory, and Cognition*, 20, 6, 1251-1263.
- Spelke, E. S., 1990, Principles of object perception, *Cognitive Science*, 14, 29-56.
- Spelke, E. S., 1994, Initial knowledge : Six suggestions, *Cognition*, 50, 431-445.
- Spelke, E. S., Breinlinger, K., Macomber, J., et Jacobsen, K., 1992, Origins of knowledge, *Psychological Review*, 99, 605-632.

- Spelke, E. S., et Tsivkin, S., 2001, Initial knowledge and conceptual change : Space and number, dans M. Bowerman et S. C. Levinson, éd., 2001.
- Sperber, D., éd., 2000, *Metarepresentations : A Multidisciplinary Perspective*, New York : Oxford University Press.
- Sperber, D., et Girotto, V., 2002, Use or misuse of the selection task ? Rejoinder to Fiddick, Cosmides, and Tooby, *Cognition*, 85, 3, 277-290.
- Sperber, D. et Wilson, D., 1986, *Relevance*, Cambridge, MA : Harvard University Press.
- Sperber, D., Cara, F., et Girotto, V., 1995, Relevance theory explains the selection task, *Cognition*, 57, 1, 31-95.
- Squire, L. R., et Knowlton, B. J., 1995, Learning about categories in the absence of memory, *Proceedings of the National Academy of Sciences of the United States of America*, 92, 12470-12474.
- Staddon, J., 2000, *The new behaviorism: Mind, mechanism and society*, Philadelphia, PA : Psychology Press.
- Stanovich, K. E., et West, R. F., 2000, Individual differences in reasoning : Implications for the rationality debate, *Behavioral and Brain Sciences*, 23, 645-726.
- Stanovich, K. E., et West, R. F., 2003, Evolutionary versus instrumental goals : How evolutionary psychology misconceives human rationality, dans D. Over, éd., *Evolution and the psychology of thinking : The debate*, Hove : Psychology Press.
- Stich, S. P., 1978, Beliefs and sub-doxastic states, *Philosophy of Science*, 45, 4, 499-518.
- Stich, S. P., 1983, *From folk psychology to cognitive science : The case against belief*, Cambridge, MA : MIT Press.
- Stich, S. P., 1990, *The Fragmentation of reason*, Cambridge, MA : MIT Press.
- Stich, S. P., 1996, *Deconstructing the mind*, Oxford : Oxford University Press.
- Stich, S. P., et Warfield, T. A., 1994, *Mental representations : A reader*, Londres : Blackwell.
- Stich, S. P., et Weinberg, J., 2001, Jackson's empirical assumptions, *Philosophy and Phenomenological Research*, 62, 3, 637-643.
- Stevens, M., 2000, The essentialist aspect of naive theories, *Cognition*, 74, 2, 149-175.
- Stevens, M., 2001, Only causation matters: Reply to Ahn et al., *Cognition*, 82, 1, 71-76.
- Thibaut, J.-P., Dupont, M., et Anselme, P., 2002, Dissociations between categorization and similarity judgments as a result of learning feature distributions, *Memory and Cognition*, 30, 4, 647-656.
- Thorpe, S. J., Delorme, A., VanRullen, R., 2001, Spike based strategies for rapid processing, *Neural Networks*, 14, 6-7, 715-726.

- Thorndike, E., 1913, *An Introduction to the theory of mental and social measurement*, New York.
- Todd, P. M., 2000, Fast and frugal heuristics for environmentally bounded minds, dans G. Gigerenzer et R. Selten, éd., *Bounded rationality, The adaptive toolbox*, Cambridge, MA : MIT Press.
- Tooby, J., et Cosmides, L., 1992, The psychological foundations of culture, dans J. Barkow, L. Cosmides, et J. Tooby, éd., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, New York: Oxford University Press.
- Tversky, A., 1977, Features of similarity, *Psychological Review*, 84, 327-352.
- Tversky, A., et Kahneman, D., 1983, Extensional versus intuitive reasoning. The conjunction fallacy in probability judgment, *Psychological Review*, 90, 293-315.
- Van Gelder, T. J., 1990, Compositionality : A connectionist variation on a classical theme, *Cognitive Science*, 14, 355-384.
- Villanueva, E., éd., 1995, *Content, Philosophical Issues vol. 6*, Atascadero, CA : Ridgeview Publishing Company.
- Villanueva, E., éd., 1998, *Concepts, Philosophical Issues, vol. 9*, Atascadero, CA : Ridgeview Publishing Company.
- Vosniadou, S., et Ortony, A., éd., 1989, *Similarity and analogical reasoning*, Cambridge : Cambridge University Press.
- Vygotsky, L. S., 1986, *Thought and language, Revised edition*, Cambridge, MA : MIT Press.
- Walker, S., 1992, Developmental changes in the representation of word-meaning : Cross cultural findings, *British Journal of Developmental Psychology*, 10, 285-299.
- Wason, P. C., et Johnson-Laird, P. N., 1972, *Psychology of reasoning : Structure and content*, Cambridge, MA : Harvard University Press.
- Weinberg, J., Nichols, S., et Stich, S. P., 2001, Normativity and epistemic intuitions. *Philosophical Topics*, 29, 1-2, 429-459.
- Whittlesea, B. W. A., 1987, Preservation of specific experiences in the representation of general knowledge, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 13, 3-17.
- Wisniewski, E. J., 1996, Construal and similarity in conceptual combination, *Journal of Memory and Language*, 35, 434-453.
- Wisniewski, E. J., 2000, Similarity, alignment, and conceptual combination : Comment on Estes and Glucksberg, *Memory and Cognition*, 28, 35-38.

- Wisniewski, E. J., 2001, On the necessity of alignment : Reply to Costello and Keane 2001, *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 27, 1, 272-277.
- Wisniewski, E. J., et Love, A., 1998, Relations versus properties in conceptual combination, *Journal of Memory and Language*, 38, 177-202.
- Wisniewski, E. J., et Medin, D. L., 1994, On the interaction of theory and data in concept learning, *Cognitive Science*, 18, 221-281.
- Wittgenstein, L., 1953, *Philosophical Investigations*, édité par G. E. M. Anscombe, Oxford : Blackwell.
- Wundt, W., 1874, *Grundzuge der physiologeschen Psychologie*, traduit par E. B. Titchener, disponible sur : <http://psychclassics.yorku.ca/Wundt/Physio/>.

## Abréviations

---

ANE : argument de la nécessité explicative

Hypothèse EN : hypothèse de l'espèce naturelle

EP : essentialisme psychologique

E<sub>m</sub>P : émergence des propriétés

HC : hypothèse de la catégorisation

HH : hypothèse de l'hétérogénéité

IF : indépendance fonctionnelle

MLT : mémoire à long terme

MT : mémoire de travail

PC : paradigme classique des concepts

PE : paradigme postulant des exemplaires

PP : paradigme postulant des prototypes

PT : paradigme postulant des théories

RM : représentation mentale

TCC : théorie composite des concepts

TP : transmission des propriétés

## Index

---

### A

- Amnésie · 12, 80-81, 187-188
- Appartenance et inclusion · 146-147, 148-149
- Apprentissage · 25, 56, 89, 153, 172, 178-179, 183-186, 187-188, 189
- Argument apparence / appartenance · 263-265
- Argument de la coordination, voir coordination
- Argument de la nécessité explicative · 14, 74-75, 83, 109, 143, 196, 226, 254-255, 281-285, 286
- Argument de la sélection des propriétés, voir sélection des propriétés
- Articulation des processus cognitifs · 131-132, 191, 243, 296-299, 300
  - Conditions déclenchantes · 191-192, 243
- Attitudes propositionnelles · 12, 19, 30-32, 41-42, 200

### C

- Catégorisation · 24, 30, 35, 38-39, 40, 41, 70, 113-114, 116, 125-126, 130, chapitre 5, *passim*, 204-205, 212-213, 231, 233, 239-240, 241-243, 248, 262-265 283
  - Définition de la catégorisation · 145-154
    - Inputs de la catégorisation · 147-150
    - Outputs de la catégorisation · 146-147
  - Hétérogénéité de la catégorisation · 144, 154, 167, 169, 193, 204
  - Modèles de la catégorisation · 154-165
- Comparaison (matching) · 156-157
- Composition des concepts · 31, 77, chapitre 6, *passim*
  - Définition de la composition · 198-204
    - et exemplaires · 217-220
    - et prototypes · 206-214
    - et théories · 214-217, 221
  - Modèle de la composition des concepts · 223

Compétence et processus cognitif · 23-24, 130-131,  
Compositionnalité · 201  
Concepts animaux · 24  
Concepts incarnés · 105, 106, 108, 295  
Concepts théoriques, voir théories  
Conceptualisme · 22  
Conditions déclenchantes, voir articulation des processus cognitifs  
Connaissances déclaratives / non-déclaratives · 80-81  
Connaissances implicites / explicites · 80-81  
Controverse sur la nature des concepts · 286-289, 292, 305  
Coordination des concepts · 267-269, 275, 299-300  
    L'argument de la coordination · 266-270  
Critère de démarcation · 26, 28, 209

## ***D***

Déduction · 24, 115-116, 133, 209  
Dilemme de l'indépendance fonctionnelle · 257-258  
Dissociation ressemblance/catégorisation · 165-171  
Diversité diachronique des concepts · 75-76

## ***E***

Effet de facilitation, voir mesure dépendante  
Effet exemplaire · 121, 180-182  
Effet prototype · 178-180, 187  
Effets de facilitation · 140, 188-189  
Efficacité causale des concepts · 37-38, 74  
Emergence des propriétés · 198, 217-218, 222  
Empirisme · 101-107  
Espèce naturelle · 60-69  
    - et essence · 64-65  
    - et loi · 64-65  
    En philosophie des sciences · 62-69  
    En psychologie · 60-61

En sémantique · 61-62  
Notion causale d'espèce naturelle · 66-69, 75, 284  
Equivalence informationnelle · 136-140  
Essentialisme psychologique · 97, 124, 126, chapitre 7, *passim*  
Exemplaire · *passim*, 91-93, 112-122, 135, 138, 176-183, 217-220, 250-251, 266-274  
Explication de l'apprentissage de catégories fondée sur la réminiscence/comparaison · 183-186  
Extraction des concepts · 136-137, 156

## **F**

Feedback extensionnel · 218-220  
Fonctions cognitives supérieures · 24

## **H**

Hétérogénéité de la catégorisation, voir catégorisation  
Hiérarchisation fonctionnelle · 76, 256-257  
Holisme · 26-29  
Hypothèse de l'hétérogénéité · *passim*, 70-74, 193-194, 224, 226, 251-252, 254-255, 274-276, 280-281, 294-299  
Hypothèse EN · *passim*, 42-45, 47-48, 59-60, 69-70, 82, 102-105, 279-280, 295

## **I**

Idéaux · 97, 100  
Indépendance fonctionnelle · 256-258  
Individuation · 5, 58, 147  
    Induction · 24, 41, 90, 114-115, 128-129, 133, 209  
    Induction et espèce naturelle · 62-64  
Inputs de la catégorisation, voir catégorisation

## **M**

Mesures de la ressemblance (linéaires / non linéaires) · 117-122  
Mesures dépendantes · 140-141, 188-191

Méthodologie habituelle · 55, 144  
Mini-théorie · 93-96  
Modèles hybrides · 135, 138-140  
Modes d'interprétation de la réalité · 246-247

## **O**

Objection méréologique · 29  
Outputs de la catégorisation, voir catégorisation

## **P**

Paradigme classique des concepts · 84-87  
Paradigme et théorie · 85, 292-293  
Paradigme postulant des exemplaires, voir exemplaire  
Paradigme postulant des prototypes, voir prototype  
Paradigme postulant des théories, voir théories  
Parcimonie · 104-105, 193-194, 254-255, 290-292  
Philosophie de l'esprit · 5, 7, 293-294  
Principe de promiscuité · 34-39, 71, 90, 141  
Principe de l'efficacité causale des concepts, voir efficacité causale  
Processus cognitif et compétence cognitive, voir compétence  
Propriétés causalement centrales · 175, 216  
Propriétés fonctionnelles · 55-56, 104, 110, 112-129, 276  
Propriétés sémantiques · 15, 18-22, 48-53, 61-62, 110, 111-112, 276  
Prototype · *passim*, 76-79, 87-91, 95, 97, 98, 106, 111, 112-122, 134-136, 137-140, 154-161, 176-191, 206-214, 249, 293-294  
Proxytype · 101-107, 294  
Psychosémantique · 5-6, 25, 51-53

## **R**

Raisonnement déductif, voir déduction  
Raisonnement inductif, voir induction  
Raisonnement probabiliste · 126, 129  
Réalisme intentionnel · 17-21

Règle · 122-129, 161-164  
Règle de la typicalité élevée · 208, 212, 214, 216, 224  
Représentationalisme · 20-21  
Ressemblance · *passim*, 27, 110, 112-124, 156-161, 165-171, 176-183, 185, 213, 248-251  
Ressemblance de famille · 160-161

## **S**

Schéma · 97-100  
Sélection des propriétés · 270-273  
Sémantique minimaliste · 25-32  
Surextension · 115, 212-213  
Symboles perceptuels · 15, 50, 84, 101, 103, 107, 295

## **T**

Temps de réaction · 136, 140  
Théorie d'arrière-plan · 93-96, 271-275  
Théorie pluraliste de la cognition · 132-133, 151, 291, 296-299  
Théories composites des concepts · *passim*, 76-78, 98-99, 193-194, 225, chapitre 8, *passim*  
Théories duelles de la cognition · 132-133, 191, 296  
Théories · *passim*, 93-96, 112, 122-129, 161-164, 214-217, 221, 245-256, 251-252, 270-274, 293-294  
Transformation, expérience de la · 169-171, 241-243  
Transmission des propriétés · 207-217  
Typicalité · *passim* 41, 89, 94, 112-117, 147, 161, 163, 176, 177-178, 209, 210-212

## **V**

Vague · 177-178

# Table des matières

---

<b>Remerciements</b>	<b>2</b>
<b>Sommaire</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>Chapitre 1 Qu'est-ce qu'une théorie des concepts ?</b>	<b>17</b>
<b>1. Qu'est-ce qu'un concept ?</b>	<b>17</b>
<i>1.1. Arrière-plan métaphysique</i>	17
1.1.1. <u>Le réalisme intentionnel</u>	17
1.1.2. <u>Représentationalisme</u>	20
1.1.3. <u>Conceptualisme</u>	22
1.2. <i>Qu'est-ce qu'un concept ?</i>	22
1.2.1. <u>La notion de concept</u>	22
1.2.2. <u>Encodage des connaissances</u>	25
1.2.3. <u>Le critère de démarcation et le problème du holisme</u>	26
1.2.4. <u>L'objection méréologique</u>	29
1.3. <i>Comparaison avec la caractérisation philosophique de la notion de concept</i>	30
1.4. <i>Conclusion</i>	32
<b>2. Hypothèses théoriques</b>	<b>32</b>
2.1. <i>Hypothèses et signification</i>	32
2.2. <i>Promiscuité et promiscuité renforcée</i>	33
2.2.1. <u>Le principe de promiscuité</u>	34
2.2.2. <u>Validité empirique du principe de promiscuité</u>	36
2.2.3. <u>Le principe de promiscuité renforcée</u>	36
2.3. <i>L'efficacité causale des concepts</i>	37
2.3.1. <u>Le principe de l'efficacité causale des concepts</u>	38

2.3.2. <u>Commentaires</u>	39
2.3.3. <u>Défense du principe de l'efficacité causale</u>	40
2.4. <i>L'hypothèse de l'espèce naturelle (l'hypothèse EN)</i>	42
2.5. <i>Conclusion : panorama</i>	44
<b>3. Qu'est-ce qu'une théorie des concepts ?</b>	<b>47</b>
3.1. <i>Quel est l'objet d'une théorie des concepts ?</i>	47
3.2. <i>Le contenu des concepts</i>	48
3.2.1. <u>Quelle sorte de connaissances est encodée par les concepts ?</u>	48
3.2.2. <u>Pourquoi les concepts possèdent-ils leurs propriétés sémantiques ?</u>	50
3.3. <i>Quelles sont les propriétés des véhicules des concepts ?</i>	54
3.4. <i>La méthodologie habituelle</i>	55
3.5. <i>Les propriétés fonctionnelles des concepts</i>	55
3.6. <i>Autres aspects des théories psychologiques des concepts</i>	56
3.7. <i>Conclusion</i>	57
<b>4. Conclusion</b>	<b>57</b>
<b>Chapitre 2 Les concepts ne forment pas une espèce naturelle</b>	<b>59</b>
<b>1. Qu'est-ce qu'une espèce naturelle ?</b>	<b>60</b>
1.1. <i>La notion psychologique d'espèce naturelle</i>	60
1.2. <i>La notion sémantique d'espèce naturelle</i>	61
1.3. <i>La notion philosophique d'espèce naturelle</i>	62
1.3.1. <u>L'idée principale</u>	62
1.3.2. <u>Espèce naturelle et essence</u>	64
1.3.3. <u>Espèce naturelle et loi</u>	65
1.3.4. <u>La notion causale d'espèce naturelle</u>	66
1.4. <i>Conclusion</i>	68
<b>2. Pourquoi les concepts ne forment-ils pas une espèce naturelle ?</b>	<b>68</b>
2.1. <i>Que signifie l'hypothèse EN ?</i>	69
2.2. <i>Comment établir que les concepts ne forment pas une espèce naturelle ?</i>	69

2.3. Pourquoi les concepts n'ont-ils pas de nombreuses propriétés en commun ?	70
2.4. L'argument de la nécessité explicative	74
2.5. Conclusion	75
<b>3. Propositions analogues</b>	<b>75</b>
3.1. Diversité diachronique	75
3.2. Les théories composites des concepts	76
3.3. L'approche de Frank Keil	78
3.4. Les représentations mentales déclaratives et non déclaratives	79
3.5. Les aspects originaux de ma position	81
<b>4. Conclusion</b>	<b>82</b>
<b>Chapitre 3 Trois types de concepts</b>	<b>83</b>
<b>1. Prototypes, exemplaires et théories</b>	<b>84</b>
1.1. Le paradigme classique des concepts (PC)	85
1.2. Le paradigme postulant des prototypes (PP)	87
1.3. Le paradigme postulant des exemplaires (PE)	91
1.4. Le paradigme postulant des théories (PT)	93
1.5. Conclusion	96
<b>2. Schémas et idéaux</b>	<b>97</b>
2.1. Variations sur les trois paradigmes	97
2.2. Théories mixtes	97
2.3. Une entité théorique originale : les idéaux	100
2.4. Conclusion	100
<b>3. Proxytypes et symboles perceptuels</b>	<b>101</b>
3.1. Les symboles perceptuels et les proxytypes	101
3.2. Un défi pour le rejet de l'hypothèse EN	102
3.3. Les concepts incarnés (« embodied concepts »)	105
3.4. Conclusion	106

<b>4. Conclusion</b>	<b>107</b>
<b>Chapitre 4 L'hétérogénéité conceptuelle</b>	<b>109</b>
<b>1. Hétérogénéité conceptuelle</b>	<b>109</b>
1.1. <i>Quand deux types de représentations mentales ont-ils peu en commun?</i>	109
1.2. <i>La prémisse 3 est révisable</i>	110
<b>2. Les propriétés sémantiques et fonctionnelles des prototypes, des exemplaires et des théories</b>	<b>111</b>
2.1. <i>Ces trois entités ont des propriétés sémantiques différentes</i>	111
2.2. <i>Les propriétés fonctionnelles : ressemblance, prototypes et exemplaires</i>	112
2.3. <i>Les propriétés fonctionnelles : calculs additif et multiplicatif de la ressemblance</i>	117
2.4. <i>Les propriétés fonctionnelles : théories et processus fondés sur les règles</i>	122
2.5. <i>Conclusion</i>	129
<b>3. La théorie pluraliste des processus cognitifs</b>	<b>129</b>
3.1. <i>Le problème</i>	129
3.2. <i>Compétence et processus</i>	130
3.3. <i>Articulation des processus cognitifs</i>	131
3.4. <i>La théorie pluraliste et les théories duelles de la cognition</i>	132
3.5. <i>Conclusion</i>	133
<b>4. L'objection de Barsalou</b>	<b>134</b>
4.1. <i>La thèse de Barsalou et ses arguments</i>	134
4.2. <i>Réfutation</i>	137
4.3. <i>Conclusion</i>	141
<b>5. Conclusion</b>	<b>141</b>
<b>Chapitre 5 L'hétérogénéité de la catégorisation</b>	<b>143</b>

<b>1. Qu'est-ce que la catégorisation ?</b>	<b>145</b>
1.1. <i>La compétence de catégorisation</i>	146
1.2. <i>Les processus de catégorisation</i>	150
1.3. <i>L'hétérogénéité de la catégorisation (HC)</i>	151
1.4. <i>Un argument pour l'hétérogénéité de la catégorisation</i>	152
1.5. <i>Conclusion</i>	153
<b>2. Les théories de la catégorisation</b>	<b>154</b>
2.1. <i>Les modèles de la catégorisation fondés sur les prototypes et sur les exemplaires</i>	154
2.2. <i>Les modèles de la catégorisation fondés sur les théories</i>	161
2.3. <i>Conclusion</i>	165
<b>3. Le processus de catégorisation et les théories : données empiriques</b>	<b>165</b>
3.1. <i>Les dissociations entre les jugements de ressemblance et les jugements de catégorisation</i>	165
3.1.1. <u>Argument</u>	165
3.1.2. <u>La pizza de Rips</u>	166
3.1.3. <u>Les expériences de transformation de Rips</u>	169
3.2. <i>Création de catégories et connaissances causales</i>	171
3.3. <i>Conclusion</i>	175
<b>4. Les processus de catégorisation, les prototypes et les exemplaires : données empiriques</b>	<b>176</b>
4.1. <i>A-t-on besoin de postuler des prototypes pour expliquer la compétence de catégorisation ?</i>	176
4.2. <i>Un argument pour l'existence de prototypes</i>	183
4.3. <i>Catégorisation et neuropsychologie</i>	186
4.4. <i>Les effets de facilitation</i>	188
4.5. <i>Conditions déclenchantes</i>	190
4.6. <i>Conclusion</i>	192
<b>5. Une objection</b>	<b>192</b>
5.1. <i>L'hétérogénéité de la catégorisation</i>	192

5.2. <i>Le problème</i>	193
5.3. <i>Réponse</i>	194
<b>6. Conclusion</b>	<b>194</b>
<b>Chapitre 6 La composition des concepts</b>	<b>196</b>
<b>1. Qu'est-ce que la composition des concepts ?</b>	<b>198</b>
1.1. <i>La compétence de composition</i>	198
1.1.1. <u>L'idée de base</u>	198
1.1.2. <u>Remarques</u>	200
1.2. <i>Qu'est-ce qu'une théorie de la composition des concepts ?</i>	202
1.3. <i>Ma thèse</i>	204
1.4. <i>Conclusion</i>	206
<b>2. Composition des concepts et prototypes</b>	<b>206</b>
2.1. <i>Les propriétés typiques des catégories complexes</i>	206
2.2. <i>La transmission des propriétés (TP)</i>	207
2.3. <i>Objection</i>	209
2.4. <i>Les jugements de typicalité relativement à des catégories complexes nouvelles</i>	210
2.5. <i>Le phénomène de la surextension</i>	212
2.6. <i>Conclusion</i>	214
<b>3. Les théories et la transmission des propriétés</b>	<b>214</b>
3.1. <i>La TP et les connaissances modales</i>	215
3.2. <i>La TP et les connaissances causales</i>	216
3.3. <i>Conclusion</i>	217
<b>4. L'émergence des propriétés (E<sub>m</sub>P)</b>	<b>217</b>
4.1. <i>Qu'est-ce que l'émergence des propriétés ?</i>	217
4.2. <i>Le feedback extensionnel : propriétés émergentes et exemplaires</i>	218
4.2.1. <u>Le feedback extensionnel</u>	218
4.2.2. <u>Feedback extensionnel et exemplaires</u>	219

4.2.3. <u>Problèmes</u>	219
4.3. <i>Propriétés émergentes et théories</i>	221
4.4. <i>Conclusion</i>	222
<b>5. Un modèle de la composition des concepts</b>	<b>222</b>
5.1. <i>Le modèle</i>	222
5.2. <i>Commentaires</i>	224
<b>6. Conclusion : la composition des concepts et la prémisse 2 de l'ANE</b>	<b>224</b>
<b>Chapitre 7 Est-ce que l'eau est H<sub>2</sub>O ?</b>	<b>226</b>
<b>1. Eau, sang et thé</b>	<b>227</b>
1.1. <i>L'essentialisme psychologique (EP)</i>	228
1.2. <i>L'essence de l'eau, c'est d'être composée de molécules de H<sub>2</sub>O</i>	230
1.3. <i>Comment utilisons-nous « eau » ?</i>	231
1.4. <i>Les critiques d'Abbott</i>	233
1.5. <i>Réponse aux critiques d'Abbott</i>	235
1.6. <i>Conclusion</i>	236
<b>2. Avons-nous un concept essentialiste d'eau ?</b>	<b>237</b>
2.1. <i>L'interprétation proposée par Malt</i>	237
2.2. <i>Données en faveur de l'interprétation de Malt</i>	238
2.3. <i>Les données contre l'interprétation de Malt</i>	240
2.4. <i>Quand ces différents concepts sont-ils activés en MT ?</i>	243
2.5. <i>Conclusion</i>	244
<b>3. EAU<sub>E</sub> et EAU<sub>NE</sub></b>	<b>245</b>
3.1. <i>L'interprétation proposée par Bloom</i>	245
3.2. <i>L'interprétation de Gelman et Hirschfeld</i>	247
3.3. <i>Quelle est la nature de EAU<sub>E</sub> et EAU<sub>NE</sub> ?</i>	248
3.4. <i>L'hypothèse de l'hétérogénéité prend-elle l'eau ?</i>	251
3.5. <i>Généralisation</i>	252

3.6. <i>Conclusion</i>	253
<b>4. Conclusion</b>	<b>253</b>
<b>Chapitre 8 Contre les théories composites des concepts</b>	<b>255</b>
<b>1. Le dilemme de l'indépendance fonctionnelle</b>	<b>256</b>
1.1. <i>L'indépendance fonctionnelle</i>	257
1.2. <i>Le dilemme</i>	258
1.3. <i>Conclusion</i>	259
<b>2. Théories composites et indépendance fonctionnelle</b>	259
2.1. <i>Les théories duelles des concepts</i>	259
2.1.1. <u>Que soutiennent les théories duelles des concepts ?</u>	259
2.1.2. <u>La critique standard</u>	261
2.1.3. <u>Limites des théories duelles</u>	261
2.2. <i>Les théories composites contemporaines</i>	262
2.3. <i>Conclusion</i>	263
<b>3. Quelle est la validité empirique de la TCC ?</b>	<b>263</b>
3.1. <i>Ressemblance et appartenance : une distinction intuitive</i>	263
3.1.1. <u>L'argument</u>	263
3.1.2. <u>Réponse</u>	264
3.2. <i>L'argument de la coordination</i>	267
3.2.1. <u>L'argument</u>	267
3.2.2. <u>Réponse</u>	268
3.2.3. <u>Coordination de nos concepts coréférentiels</u>	270
3.3. <i>L'argument de la sélection des propriétés</i>	271
3.3.1. <u>L'argument</u>	272
3.3.2. <u>Réponse</u>	275
3.4. <i>Conclusion</i>	275
<b>4. La seconde branche du dilemme</b>	<b>275</b>

4.1. Deux variantes terminologiques	276
4.2. Le « bon niveau » de la recherche psychologique	277
4.3. Quelle variante terminologique préférer ?	278
4.4. Conclusion	278
<b>5. Conclusion</b>	<b>278</b>
<b>Conclusion</b>	<b>280</b>
<b>1. Les concepts ne forment pas une espèce naturelle</b>	<b>280</b>
1.1. L'hypothèse EN	280
1.2. Le rejet de l'hypothèse EN	281
1.3. L'argument de la nécessité explicative	282
1.4. Remarques	284
1.5. Conclusion	286
<b>2. Mettre fin à la controverse sur la nature des concepts</b>	<b>286</b>
2.1. La controverse sur la nature des concepts	286
2.2. De l'utilité des fausses théories	288
2.3. Les principes d'unification et de parcimonie	290
2.4. Conclusion	292
<b>3. De nouveaux thèmes de recherche</b>	<b>293</b>
3.1. Philosophie des concepts : de la philosophie de l'esprit à la philosophie de la psychologie	293
3.2. Questions non résolues	294
3.3. Nouveaux thèmes en psychologie : la théorie pluraliste de la cognition	296
3.4. Nouveaux thèmes en psychologie : la question de la coordination et l'évolution des concepts	299
3.5. Conclusion	300
<b>4. L'avenir des concepts</b>	<b>300</b>
4.1. Le problème	300

<i>4.2. Un problème mineur</i>	301
<i>4.3. Eliminer ou conserver ?</i>	301
<i>4.4. Avons-nous des concepts ?</i>	305
<b>5. Conclusion</b>	<b>305</b>
<b>Bibliographie</b>	<b>307</b>
<b>Abréviations</b>	<b>333</b>
<b>Index</b>	<b>334</b>
<b>Tables des matières</b>	<b>336</b>