



HAL
open science

Concept et changement de concept : Concept, contenu et inférence, Bases pour une approche dynamique du concept.

Benjamin Sylvand

► To cite this version:

Benjamin Sylvand. Concept et changement de concept : Concept, contenu et inférence, Bases pour une approche dynamique du concept.. domain_scco.cogdynamics. Université Paris-Sorbonne - Paris IV, 2005. Français. NNT : . ijn_00000655

HAL Id: ijn_00000655

https://theses.hal.science/ijn_00000655

Submitted on 15 Dec 2005

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 V “CONCEPTS ET LANGAGES”

N°

THÈSE

pour obtenir le grade de
DOCTEUR DE L’UNIVERSITÉ PARIS IV
Discipline : Philosophie

présentée publiquement et soutenue par

Benjamin Éric Quasar SYLVAND

le 10 Décembre 2005

Concept et Changement de concept
Concept, contenu et inférence
Bases pour une approche dynamique du concept

Directeur de thèse :

Prof. Pascal ENGEL

JURY

Mr. Jacques DUBUCS, *IHPST, Université Paris I – UMR 8590 CNRS*

Mr. Jérôme DOKIC, *EHESS*

Prof. Pierre LIVET, *Université de Provence, Aix-en-Provence*

Prof. Göran SUNDHOLM, *Université de Leiden*

Concepts et Changement de concept
Bases pour une approche dynamique du concept

Benjamin SYLVAND

2005

à Pascal Engel

« Un couteau sans lame auquel il manque le manche. »

[Georg Christoph Lichtenberg, *Le couteau sans lame* Collection romantique n° 72 José Corti, 1999, p. 54]

«Non ci sono indizi, né prove, né motivazioni. Siamo di fronte a un omicidio ma non c'è il morto, non c'è nemmeno il movente e manca l'arma del delitto.»

[«C'est comme si je me trouvais dans un procès pour homicide sans qu'il y ait de cadavre, d'arme du délit ni de mobile.»]

[Silvio Berlusconi, *Libération*, 18 – 06 – 2003]

«Quand les jumeaux demandèrent à quoi servaient les boutons de manchette et s'entendirent répondre, par Ammu, que c'était pour «boutonner les manches», pareille logique linguistique, dans ce qui, jusqu'ici, leur avait paru être une langue illogique, les réjouit au plus au point. *bouton + manchette = bouton de manchette*. Seules la rigueur et la logique des mathématiques pouvaient prétendre rivaliser avec un tel phénomène.»

[Arundhati Roy, *Le Dieu des Petits Riens*, Folio Gallimard, 1998, p. 81]

«Lorsqu'il retira son masque, en sortant dans l'air frais du matin, il se rendit compte qu'il venait d'absorber une dose massive d'un lointain recoin de la réalité dont une grande majorité du pays ne soupçonnait même pas l'existence. Cela jeta une lumière entièrement neuve sur le concept de dinde de Noël garnissant un rayonnage de supermarché.»

[Tristan Egolf, *Le seigneur des porcheries*, Folio, Gallimard, p. 218]

« Du palais royal je ne dirais rien, parce que nous ne l'avons aperçu que de loin et n'avons appris que ce que les Chinois nous en ont dit, choses si extraordinaires qu'il est à craindre de les rapporter. C'est pourquoi je n'en traiterai pas pour l'instant, car je dois plus avant faire le récit de ce que nous vîmes dans celui de la ville de Pékin, et j'avoue redouter déjà d'avoir à rendre compte du peu même que nous en observâmes. Non que cela puisse paraître étrange à qui a vu les autres merveilles de ce royaume de Chine, mais parce que je crains que ceux qui voudront mesurer l'abondance de ce qui se trouve dans les contrées qu'ils n'ont point vues à l'aune du peu qu'ils voient dans les contrées où ils ont grandi ne veuillent opposer le doute – ou peut-être refuser tout crédit – à ces choses, qui ne s'accordent pas à leur entendement et à leur pauvre expérience.»

[Fernão Mendes Pinto, *Pérégrination*, Minos, La Différence, 2002, 88, pp. 327-328]

Table des matières

Liste des tableaux	11
Liste des figures	13
I Avant-Propos	15
Remerciements	17
Remarques Typographiques	19
II Problème	23
1 Présentation et portée du problème	25
2 Analyse du problème	35
2.1 Contenu non-conceptuel et contenu conceptuel	35
2.2 Différence entre le non-conceptuel et le conceptuel	37
2.2.1 Représentation	37
2.2.2 Caractère épistémique	41
2.2.2.1 Caractère épistémique fort	41
2.2.2.2 Caractère épistémique faible	41
2.2.2.3 Caractère non-épistémique fort	42
2.2.2.4 Caractère non-épistémique faible	43
2.2.3 Sens/Référence	44
2.3 Concept	45
2.3.1 Individuation du concept	45
2.3.2 Catégoricité du concept	48
2.3.3 Combinaison de concepts	54
2.3.4 Apprentissage du concept	59
2.3.5 Publicité du concept	64

3	Conceptions du concept	69
3.1	Théorie classique	69
3.2	Théorie-théorie	75
3.3	Exemplaires	82
3.4	Prototypes	87
3.5	Empirisme conceptuel	98
3.6	Atomisme conceptuel	104
3.7	Rôles conceptuels	111
III	Bases pour une approche dynamique du concept	119
4	Liminaires	121
5	Inférence	141
5.1	La distinction des concepts	145
5.2	L'inférence comme maîtrise du concept	147
5.3	Inférence et compositionnalité	155
5.4	Inférence et théorie des concepts	162
6	Changement conceptuel	181
6.1	Raisons pour de changer de concept	184
6.1.1	De la non pertinence avérée du concept	184
6.1.2	Deux concepts coréférentiels différents entraînent la révision ou l'abandon de l'un des deux au moins.	186
6.1.3	L'affaiblissement ou le renforcement des raisons ga- rantissant l'usage du concept entraîne la révision de celui-ci	190
6.1.4	La modification de l'extension du concept modifie celui-ci	192
6.1.5	Toute suspicion sur l'usage ou l'application du concept doit conduire à sa révision	193
6.2	Réquisits d'un changement de concept	195
6.2.1	Les concepts d'un agent doivent être consistants	199
6.2.2	Les concepts d'un agent doivent être clos inférentiellement	200
6.2.3	Principe de remise en cause positive	201

6.2.4	Principe de remise en cause négative	202
6.2.5	Principe de préférence dans le changement	203
6.2.6	Principe de minimalisation dans le changement	204
6.2.7	Principe de correspondance catégorielle	205
7	Concept et croyance	209
7.1	Expansion	209
7.1.1	Expansion de croyance	209
7.1.2	Expansion de concept	210
7.1.2.1	Expansion de l'application du concept	210
7.1.2.2	Expansion de l'usage du concept	210
7.1.2.3	Pourquoi parler de changement de concept plutôt que de remplacement ?	212
7.2	Contraction	213
7.2.1	Contraction de croyance	213
7.2.2	Contraction de concept	213
7.2.2.1	Contraction d'application de concept	214
7.2.2.2	Contraction d'usage du concept	215
7.3	Révision	215
7.3.1	Révision de croyance	215
7.3.2	Révision de concept	216
7.3.2.1	Révision de l'application du concept	216
7.3.2.2	Révision d'usage du concept	217
7.3.2.3	Changement de conception	218
8	Conclusions et perspectives	221
8.1	Conclusions	221
8.2	Perspectives	227
IV	Annexes	229
9	Connaissance et incertitude.	231
10	Tracking Agency	239
10.0.1	Abstract	239
10.1	Individuation : Creating and Maintaining Reference	240
10.1.1	Reference to Objects <i>vs.</i> Reference to Agents in Infants	240

10.2	Perceptual Tracking of Individuals	243
10.2.1	Object Files and Their Dynamics	244
10.2.2	From Object Files to Infants' Representation of Objects	245
10.2.2.1	Proto-objects vs. commonsense objects	245
10.3	Perceptual Tracking of Agency	246
10.3.1	Extending the File Notion to Agents	248
10.3.1.1	Agent Files vs. Agency Cues	249
10.4	Object vs. Agent Tracking : Empirical Directions	250
10.4.1	Object vs. Agent File Fixation Cues	250
10.4.2	Object vs. Agent File Preserving Properties	252
10.4.2.1	An Objectual Bias in Agent Files Preserva- tion ?	253
10.4.3	Object vs. Agent File Content	254
10.5	Are There Really Agent Tracking Mechanisms?	255
10.6	Conclusions	257
11	Percept	259
11.0.1	Abstract	259
11.1	Introduction	259
11.2	The conceptual and non-conceptual contents	260
11.2.1	The epistemic and non-epistemic levels	262
11.2.2	The representational aspect	265
11.2.3	Proto-judgments	267
11.3	The use of proto-judgments in explaining the Müller-Lyer illu- sion	270
11.3.1	The theoretical levels of experience	273
11.4	Conclusion	275
	Bibliographie	279
	Index	302

Liste des tableaux

3.1	Table de vérité de la conjonction	94
5.1	Table de vérité de la conjonction	165

Table des figures

2.1	Illusion de constance de taille (image : David Falco)	35
2.2	Constance de taille (image : David Falco)	36
2.3	Contraste dans l'illusion de constance de taille (image : David Falco)	37
8.1	Schéma récapitulatif de la structure du niveau conceptuel et du niveau non-conceptuel	228
10.1	Playing 'footsie' and agent individuation	247
10.2	The Asteroids Game : objects vs. agents	251
11.1	Müller-Lyer illusion	271

Première partie

Avant-Propos

Remerciements

Le travail présenté ici est une thèse de philosophie. Il ne faut pas le considérer comme une œuvre de philosophie, mais comme un exercice d'école. Je le considère comme un travail autonome, mais aussi dans la continuité de mes mémoires de maîtrise – « La querelle des idées abstraites entre Locke et Leibniz et ses prolongements dans le débat contemporain » – et de DEA (Diplôme d'Études Approfondies) – « Le problème des abstraits mathématiques chez Frege, Husserl et dans la philosophie de l'arithmétique contemporaine », tous deux présentés à l'université Paris 4 Sorbonne. Cette thèse achevant un cycle de formation en philosophie, j'aimerais en profiter pour remercier tous ceux qui m'auront initié à cette discipline et me l'auront enseigné.

J'ai bénéficié durant les années 2001 à 2004 d'une Allocation de Recherche de la part du Ministère Français de la Recherche, ce qui m'a permis de financer durant ces années cette recherche. Cette allocation était couplée à un Monitorat à l'Université Paris 4, ce qui m'a permis de m'initier à l'enseignement. Je voudrais remercier ici ces institutions, mes collègues et les étudiants qui auront suivi mes enseignements et de qui j'ai beaucoup appris en retour.

Depuis 2001, l'Institut Jean Nicod (CNRS/EHESS/ENS) m'a accueilli parmi ses étudiants. J'avoue avoir trouvé là une atmosphère et des conditions de travail sans lesquelles ce travail n'aurait jamais les formes qu'il a actuellement, et peut-être resterait-il une thèse sans sujet ni chapitre. Ces années passées à l'institut furent capitales dans ma formation, notamment en sciences cognitives, et les divers séminaires, universités d'été, conférences et autres modes de collaboration ont été autant de possibilités de partage d'idées extrêmement stimulants. Je tiens à remercier chaleureusement

l'ensemble des membres de l'institut.

Je ne compte pas le nombre de discussions et d'échanges divers que j'ai pu avoir durant ces années, je voudrais remercier tous ces interlocuteurs, et en particuliers mes co-auteurs et collaborateurs à la *European Review of Philosophy* : Dario Taraborelli, Nausicaa Pouscoulous, Christine Van Geen, Àngeles Eraña, Nivedita Gangopadhyay, Paul Égré, Nicolas Bullot, Damian Justo, Julien Dutant, Benoît Hardy-Vallée, David Nicolas.

Merci également à David Falco pour ses images.

Je tiens à remercier l'ensemble des membres du jury qui ont lu et examiné ce travail : Göran Sundholm, Pierre Livet, Jacques Dubucs, et Pierre Jacob. Merci aussi à Élisabeth Pacherie pour son avis sur une version préliminaire de ce travail. Merci à Jérôme Dokic et Pierre Livet d'avoir rédigé les pré-rapports de ce travail.

S'il me fallait formaliser ma dette envers Pascal Engel, je serais obligé d'employer un quantificateur universel non restreint, tant je lui dois. Je ne sais comment le remercier. Ce travail lui est dédié.

Remarques typographiques

Le symbolisme utilisé a vocation de faciliter la lecture. Dans la mesure du possible toute introduction de signe logique a été expliquée et toute formule paraphrasée, de sorte que le texte devrait se suffire à lui-même.

Typographie

Les PETITES CAPITALES désignent des concepts.

Les *italiques* désignent des propriétés.

Les ‘guillemets simples’ encadrent des termes lexicaux.

Symboles

- \neg ***négation*** : inverse la valeur de vérité de la formule préfixée de ce signe.
- \wedge ***conjonction*** : la formule est vraie lors que le terme immédiatement à gauche et le terme immédiatement à droite de ce signe sont vrais.
- \vee ***disjonction inclusive*** : correspond au ‘ou’ de la langue française. La formule est vraie si au moins un des termes situés immédiatement à gauche ou immédiatement à droite du signe est vrai, fausse sinon.
- \rightarrow ***implication*** : la formule est vraie si le signe situé à droite de la flèche est vrai ou si le terme à gauche et le terme à droite du signe sont tous les deux faux.
- \leftrightarrow ***biconditionnel*** : la formule est vraie si et seulement si les termes à gauche et à droite du signe ont la même valeur de vérité.
- \in ***inclusion*** : le terme situé à gauche du signe appartient au terme situé à droite du signe.
- ***barre de jugement*** : la formule préfixée de ce signe est un jugement.

- \vdash *barre d'assertion* : la formule préfixée de ce signe est un jugement tenu pour vrai.
- \dashv *barre de dénégation* : la formule préfixée de ce signe est un jugement tenu pour faux.
- \models *barre de tautologie* : la formule préfixée de ce signe est logiquement vraie.
- \square *nécessité* : la formule préfixée de ce signe est vraie dans tous les modèles.
- \diamond *possibilité* : la formule préfixée de ce signe est vraie dans au moins un modèle.
- \cap *intersection ensembliste* : les éléments communs aux ensembles situés à droite et à gauche du signe forment un ensemble.
- \cup *union ensembliste* : les éléments des ensembles situés à gauche et à droite du signe forment un ensemble.
- $x, y, z \dots$ *variable* : le symbole peut être substitué dans la formule par n'importe quelle constante d'ensemble qu'elle représente.
- $a, b, c \dots$ *constante* : le symbole vaut pour un individu.
- A, B, C *constante de prédicat* : le symbole vaut pour un prédicat.
- $\varphi, \psi \dots$ *formule* : le symbole vaut pour une formule.
- λ *lambda* : la variable liée par ce symbole est abstraite de la formule dans laquelle elle apparaît.
- ϵ *opérateur de Hilbert* : la formule préfixée de ce symbole a une extension.
- ι *opérateur de Russell* : la variable préfixée de ce symbole ne vaut que pour un seul individu.
- α, β, \dots *constante de concept* : le symbole vaut pour un concept.
- $\lceil \rceil$ *indication de concept* : la formule entre ces crochets est un concept.
- a_1, a_n *indice* : numération de la constante indiquée.
- Γ, Δ, \dots *constante de jugement* : le symbole vaut pour un jugement.
- $\dot{=}$ *compatibilité* : le terme à gauche du signe est compatible avec le terme à droite du signe.
- $\dot{\neq}$ *incompatibilité* : le terme à gauche du signe n'est pas compatible avec le terme à droite du signe.

-
- \exists *quantificateur existentiel* : un individu au moins substitué à la variable liée par ce symbole vérifie la formule préfixée.
- \forall *quantificateur universel* : tous les individus substitués à la variable liée vérifient la formule.
- \perp *contradiction* : la formule est simultanément vraie et fausse.
- $\dot{\vdash}$ *dérivation conceptuelle* : la formule à droite du signe est un concept dérivé de la formule située à gauche du signe.
- $\dot{\nmid}$ *non dérivation conceptuelle* : la formule à droite du signe est un concept qui ne peut pas être dérivé de la formule située à gauche du signe.
- $\dot{\in}$ *inclusion conceptuelle* : le concept situé à gauche du signe est compris dans le concept situé à droite du signe.
- $\dot{\notin}$ *exclusion conceptuelle* : le concept situé à gauche du signe n'est pas inclus dans le concept situé à droite du signe.
- * *révision* : la formule exposée de ce signe est révisée.
- $\dot{-}$ *élimination conceptuelle* : la formule située à gauche du signe est retirée de la formule située à droite du signe.
- \sqcup *variable non-conceptuel* : le symbole indique un contenu non-conceptuel.
- \mathcal{F} *fonction* : le symbole vaut pour une fonction.
- \mathfrak{f} *fonction de correction* : le symbole vaut pour une fonction de correction.
- \mathcal{M} *modèle* : le symbole désigne le modèle d'évaluation d'une formule.

Ponctuation

Là où il a été jugé nécessaire de clarifier le texte, des symboles de ponctuation ont été introduits.

- () *parenthèses* : les termes à l'intérieur des parenthèses forment une formule. Toujours symétriques.
- [] *crochets* : remplacent les parenthèses lorsque la profondeur est égale ou supérieure à deux. Toujours symétriques.
- { } *accolades* : les termes à l'intérieur des accolades forment un ensemble. Toujours symétriques.

, *virgule* : dans les formules, séparent les termes listés.

< > *crochets pointus* : les termes entre ces signes sont ordonnés. Toujours symétriques.

Citations

Les citations ont été laissées en langue originale dans le corps du texte et traduites en note. La traduction est personnelle et seul l'original est tient lieu de référence.

Deuxième partie

Problème

Chapitre 1

Présentation et portée du problème

La question de savoir ce qu'est un concept court à peu près tout le long de la philosophie. Essayer de faire une histoire de cette notion revient plus ou moins à faire une histoire de la philosophie. Comprendre ce qu'est un concept c'est tout simplement vouloir comprendre comment il se fait que je pense que l'objet qui est devant moi est une tasse lorsque c'est une tasse, et que vous – qui êtes différents de moi – pensiez aussi que c'en est une, si bien que nous pouvons communiquer, échanger quelque chose à propos du monde. En bref : *comment un agent cognitif peut penser à quelque chose à propos du monde ?* Cette question paraît naïve et pas bien différente de celles que posent les enfants et qui agacent les adultes. C'est pourtant à cette question que cette étude s'attaque. Plus spécifiquement encore, la question centrale de ce travail est le changement de concept. Comment un agent cognitif qui possède un concept peut le modifier en fonction de l'expérience et d'informations nouvelles qu'il pourrait avoir.

Mais bien sûr, une expédition, aussi bien préparée soit elle, se révèle toujours être différente que prévue, sur le terrain. Le plan initial de cette recherche était de comprendre qu'elle était la structure logique du concept, étant entendu que celui est l'entité mentale qui permet à un agent cognitif qui la possède de pouvoir penser à plusieurs éléments différents possédant une même propriété. Cet angle d'attaque s'inspirait de recherches précédemment menées, d'une part sur les idées abstraites et la querelle

à leur propos chez Locke, Leibniz et ses prolongements dans le débat contemporain, notamment chez Putnam et Goodman, et une étude sur la fonction des entités abstraites dans les mathématiques, en partant de Frege, Husserl jusqu'à la philosophie de l'arithmétique contemporaine. Croisant ces deux études, le concept devait être quelque chose comme une entité abstraite ayant la même fonction qu'un terme générique, mais restait à savoir comment cela se passait dans l'esprit.

Les études sur le concept sont très nombreuses depuis une trentaine d'années, motivées par la linguistique d'une part, les avancées de l'intelligence artificielles et de la psychologie d'autre part. Ces terrains de recherches n'ont pas toujours travaillé de concert, mais s'ils sont regroupés, parfois un peu artificiellement, sous le label « science cognitives ». Les données dans tous ces domaines sont si nombreuses et variées qu'il aurait été impossible même d'espérer de les traiter toutes. Ma formation étant philosophique, l'axe choisi est philosophique. Il se fixe particulièrement autour de trois auteurs : Jerry Fodor, auteur du *Concepts, Where Cognitive Sciences Went Wrong* de 1998, qui a fait grand bruit, Jesse Prinz, avec son *Furnishing the Mind : Concept and Their Perceptual Basis* de 2002, qui est une référence qu'il faut discuter même si son ouvrage est imparfait sous bien des aspects, et le *A Study of Concepts* de 1992, de Christopher Peacocke, qui reste sans doute l'ouvrage le plus aboutit sur la question des concepts. Cette présente étude s'en inspire beaucoup et en reprend l'approche normative et théorique.

Mais comme les expéditions polaires, alors que l'on planifiait un périple de cent cinquante kilomètres, l'avance sur le terrain est d'une petite dizaine seulement. Ce qui pourrait se prendre comme un échec sur le papier, peut cependant se révéler suffisant dans la pratique. L'objet de ce travail a donc évolué durant ces quatre années, et au final c'est un problème dans le problème qui se retrouve au centre de l'analyse : le changement de concept. Comment un agent cognitif possédant un concept peut adapter celui-ci à son expérience ou à ses autres concepts ? Que devient par exemple le concept BALEINE que Paul possédait lorsqu'il pensait que les baleines étaient des poissons et qu'il sait maintenant que ce sont des mammifères ? Qu'est-ce qui a changé dans la représentation du monde du fils de Carolyn qui ne nommait 'canards' tous les oiseaux aquatiques et qui maintenant

appelle certains ‘hérons’ et d’autres ‘sternes’? Est-ce que Pascal a modifié quelque chose dans sa manière de voir le monde lorsque Pierre lui apprend les règles d’arbitrage du sabre? Quels changements s’opèrent dans les agents cognitifs lorsqu’ils passent d’une cosmogonie géocentrique à une cosmogonie héliocentrique? La dame qui se plaignait d’arthrite en se frottant la cuisse, a-t-elle toujours mal à la cuisse ou doit-elle avoir des douleurs dans le genou maintenant que le médecin lui a dit que l’arthrite était une inflammation articulaire?

Évidemment, pour pouvoir examiner la question du changement de concept et savoir s’il est possible de parler de changement au niveau du concept – le changement impliquant une dynamique et la conservation de certains éléments au sein du concept alors que d’autres sont modifiés selon certaines conditions – ou bien s’il n’y a qu’un remplacement de concept par un autre lors d’un changement de situation, il faut savoir ce qu’est un concept.

La première partie de cette étude, « [Analyse du problème](#) », porte sur la notion de concept et sur son domaine, en particulier de ce qui distingue le concept de ce qui ne l’est pas, en d’autres termes, le *non-conceptuel*. De cette distinction de domaines, se dégagent les caractéristiques de ce qu’un concept doit avoir pour être un concept. À partir de cela, dans la section « [Littérature sur le concept](#) », je présente brièvement et discute les positions les plus saillantes et dominantes des trois dernières décennies sur les concepts. Cette discussion n’est pas exhaustive et, même si elle se veut franche et honnête est certainement caricaturale. La principale conclusion on s’en doute est qu’aucune de ces théories en parvient réellement à rendre compte du problème du changement de concept.

De fait, dans une seconde partie, « [Bases pour une approche dynamique du concept](#) », en particulier dans le chapitre « [Liminaires](#) » je propose une structure pour comprendre plus finement l’entité mentale qui permet à un agent qui la possède à la fois de désigner des objets de l’environnement dans lequel il évolue, et de mettre en relations ces représentations entre elles. Le concept se retrouve être une entité complexe dotée d’une extension – une application à un objet différent du concept – et d’une intension – des liaisons inférentielles avec d’autres concepts ou entités mentales. De plus, il faut supposer que l’agent possède une sorte de théorie cognitive

naïve, qu'il soit en mesure de penser le point de vue d'un autre agent sur l'objet qu'il représente. Par exemple, face à un animal, je peux penser que l'objet en question est un ours polaire, dans ce cas j'applique le concept OURS POLAIRE à cet objet, je peux être en mesure de me dire que si l'objet en question fait tel ou tel mouvement ou a telle réaction alors je vais tirer une salve de semonce, dans ce cas je relie inférentiellement le concept OURS POLAIRE au concept DANGER, et je peux me dire qu'un biologiste par exemple pourrait corriger ma représentation en me disant que ce n'est pas un ours polaire mais un renne par exemple. Mais cela implique d'une manière ou d'une autre que je me pose la question de ce qu'un scientifique penserait de l'objet que j'ai en face de moi. Cela paraît être une contrainte très forte. Elle l'est. Mais cela ne signifie pas que pour chaque représentation que l'agent a du monde, il se pose effectivement la question. La contrainte dit simplement qu'il doit pouvoir le faire. C'est là l'un des traits fondamentaux de la théorie des concepts défendue ici, et c'est ce qui distingue le concept des niveaux non-conceptuels du traitement de l'information qui provient de l'expérience, et ce qui distingue aussi la théorie des concepts défendue ici des autres théories des concepts, en particulier de celles exposées dans la première partie. Cette analyse du concept est théorique et *a priori*, c'est ce qui en fait sa faiblesse sans doute – et pourquoi les psychologues se demandent à quoi elle peut servir – mais également ce qui en fait sa force, dans la mesure où elle permet de faire un certain nombre de prédictions, qu'il ne serait sans doute pas inintéressant de valider ou d'infirmier expérimentalement. Mais c'est là un prolongement qui reste à faire. Chaque chose en son temps.

La structure complexe du concept exposée dans les « [Liminaires](#) » de la seconde partie, repose sur la notion d'*inférence* et de l'importance accordée aux relations entre les concepts. L'*inférence* est une certaine relation entre des entités compatibles, c'est une notion logique qu'il faut expliquer et examiner plus avant, afin de comprendre comment sa structure implique certaines contraintes sur les entités auxquelles elle s'applique. En simplifiant beaucoup on pourrait dire que les théories des concepts se soucient d'abord peu ou pas du tout de la notion de non-conceptuel et ne cherchent pas à différencier ces deux notions, mais également qu'elles considèrent que le seul but ou du moins la vocation principale du concept est de catégoriser

le monde. Ainsi la théorie du concept devient une théorie de la classification. Il y a deux manières, pour le dire rapidement, de ranger une chambre : l'une consiste à prendre des boîtes sur lesquelles on écrit ce qu'elles doivent contenir et on cherche dans le bazar s'il y a quelque chose qui entre dans la boîte, ou bien on fait des tas à partir des jouets éparpillés en fonction de ressemblances qu'ils peuvent plus ou moins avoir entre eux, et lorsque les tas sont suffisamment gros, on les met dans des boîtes. La première méthode est intensionnelle, la seconde extensionnelle. Il m'a toujours semblé surprenant que suivant la méthode choisie, la chambre n'était pas du tout rangée de la même manière. En un sens cela semble évident, mais d'un point de vue logique, on pourrait se dire qu'en définitive, comme il s'agit des mêmes objets et de la même personne qui les range, le résultat devrait être plus ou moins le même. Et bien non. Pire, les deux méthodes ne sont pas compatibles sur la totalité de leur développement, et pourtant pour finir de ranger la chambre, c'est-à-dire, pour la ranger complètement, il faut faire appel aux deux. Ce sont ces derniers objets dont on ne sait jamais quoi faire et qu'on met arbitrairement dans une boîte ou dans une autre à la fin du rangement. Cela s'explique logiquement, c'est ce que montre le théorème de Ramsey – ce résultat de logique combinatoire établie simplement que si l'on décide de ranger un certain nombre d'objet dans un nombre plus petit de boîtes, alors forcément plus au moins deux objets se retrouveront dans une même boîte, ce résultat est naïvement trivial, mais logiquement il dit que pour tout coloriage de n -uplets d'entiers en un nombre fini de couleur, il existe un ensemble infini homogène et cela cause un problème lorsque dans une classification on remplace les couleurs par des labels et les entiers par des objets, cela crée des boucles qui empêche de retrouver certains objets pourtant classés.

Or si le concept nous permet effectivement de classer les éléments du monde, il ne fait pas que cela, parce que lorsqu'on met les derniers objets qui ne ressemblent à rien dans une boîte au hasard en se disant que de toute manière il y a peu de chance qu'on en ait besoin tout de suite, on possède bien un concept de ces choses-là. Tout ce qui nous apparaît en fait est déjà conceptualisé. La manière d'expliquer des deux aspects du concept, le fait qu'il s'applique à un objet et qu'il est relié à d'autres, est de rendre compte de son aspect intensionnel. L'aspect intensionnel du concept est l'ensemble des relations qu'il entretient avec d'autres concepts. De sorte que les concepts forment des trames, des *conceptions*, sur lesquelles se développent

les croyances, les connaissances, les jugements, les désirs, les plans d'actions et tout autre état mental d'ordre supérieur. Cela implique que la théorie des concepts soutenue ici est holistique, du moins partiellement. Les connexions au niveau de la trame sont inférentielles, ce qui explique le chapitre « [Inférence](#) » consacré à cette question.

Le point de focal de l'étude du concept mené ici est le *changement de concept*. Pourquoi avoir choisi cet axe ? Parce qu'il suppose de comprendre non seulement ce qu'est un concept, mais également comment un agent cognitif le possède et l'apprend, et aussi comment les agents peuvent partager et communiquer des représentations mentales. C'est donc un angle relativement complet sur le concept. C'est aussi une question qui n'est pas ou que très peu abordée d'un point de vue philosophique. Il existe des études, notamment le livre de Susan Carey, *Conceptual Change in Childhood* de 1987 sur la question, mais généralement le concept est considéré de manière statique, et il est surprenant de ne pas trouver, ne serait-ce dans les trois ouvrages de Fodor, Prinz et de Peacocke, précédemment cités, de théorie précise sur la question. Quant aux études sur les changements de concepts théoriques, telles que celles de Thomas Kuhn par exemple, elles ne relèvent pas de la question des concepts au sens entendu ici, mais du remplacement de termes théoriques au sein de théories scientifiques. Je prétends que ces changements s'opèrent au niveau des conceptions et non pas au niveau des concepts eux-mêmes. Quelqu'un m'objectait lors d'une conférence qu'une théorie du changement de concept devait trancher la question de la nature corpusculaire ou ondulatoire de la lumière, au sens où il y a deux concepts de lumière pour un même phénomène. Je continue à penser que dans ce cas nous avons deux concepts mutuellement exclusifs décrivant un même phénomène suivant des points de vue différant. En appliquant l'un ou l'autre nous ne changeons pas le concept LUMIÈRE mais nous changeons *de* concepts, nous passons du concept LUMIÈRE_{corpuscule} au concept LUMIÈRE_{onde} comme on passe du concept OURS POLAIRE au concept RENNE lorsqu'on s'est trompé dans l'identification d'un objet du monde. Lorsqu'il est question ici de changement de concept, il est question de changement *au sein d'un même concept*. Par exemple, est-ce que le fait de se retrouver dans un environnement où vous devez porter un fusil 7,62 et constamment scruter à la jumelle le paysage pour éviter les ours

polaires, ajoute quelque chose à votre concept OURS POLAIRE. Je pense que oui. Je pense que quelqu'un qui a déjà eu en face de lui un ours polaire n'a pas le *même* concept que celui qui n'a pas fait cette expérience. Or nous parlerons bien de la même chose, nous utiliserons bien de la même manière le même concept, mais n'en aurons pas le même contenu. La différence des contenus liés à l'expérience, dépend de ce que j'appelle ici le *contenu cognitif du concept*. Mais lorsque nous communiquons, lorsque nous parlons par exemple des ours polaires, nous discutons en *postulant* – c'est là qu'intervient la théorie naïve de l'esprit – que notre interlocuteur entend la même chose que nous par ce concept, ou par le terme lexical lié à celui-ci. Lorsque je parle des ours polaires je ne m'imagine pas que vous ayez déjà évolué au Svalbard avec un fusil à l'épaule, et que même si vous n'avez pas croisé d'ours, le simple fait de devoir le craindre ait ajouté quelque chose à votre concept. Je m'imagine que vous possédez au moins le *concept commun* d'ours polaire, ce que j'appelle ici le *contenu canonique* du concept. Contenu qui se situe au niveau du sens commun.

L'idée défendue ici est que le concept peut évoluer. Que je peux avoir un concept d'OURS POLAIRE qui peut évoluer dans le temps, suivant l'expérience et la connaissance que je peux avoir des ours polaires. Dire que le concept change c'est dire que quelque chose reste commun à la nouvelle version du concept et à sa ou ses versions précédentes. Par exemple je peux savoir toujours la même chose biologiquement de l'ours polaire, mais ne pas appliquer ce concept aux mêmes objets du monde, ou au contraire l'appliquer aux mêmes objets mais changer mes croyances ou connaissances à son sujet. Dans le chapitre « [Changement conceptuel](#) » j'examine ce qui peut conduire un agent cognitif à changer de concept, qu'elles sont les raisons qui déterminent ou devraient déterminer ce changement, et comment ce changement doit ou devrait se faire pour parler de changement au sein du concept et non pas de changement *de* concepts, c'est-à-dire du remplacement d'un concept par un autre.

La distinction entre le *concept* et la *conception*, c'est-à-dire entre l'entité isolée qui permet à un agent qui la possède de représenter le monde d'une part, et la trame formée par l'ensemble ou une partie des concepts possédés par cet agent, implique d'expliquer les différences entre ces deux niveaux,

au moins sur la question du changement. L'idée suivie dans le chapitre « [Concept et croyance](#) » est de prendre comme cas paradigmatique de conception, les croyances, et d'expliquer en quoi le changement de croyance est différent du changement de concept. Si le concept est considéré comme l'entité mentale qui permet de constituer des états mentaux d'ordres supérieurs, que la croyance se situe à un niveau supérieur que le concept, le changement de croyance doit être plus spécifique que le changement de concept.

Il reste à savoir s'il est possible d'isoler le concept et de l'étudier indépendamment de toutes ses applications ou utilisations. Si les croyances, les connaissances, les jugements, les désirs, les plans d'actions et autres états mentaux sont constitués de concepts, c'est que leurs études et analyses requièrent nécessairement de se pencher sur le concept. Ils en sont en quelque sorte dépendants. Cependant, le concept est-il indépendant de ces états mentaux ? Un agent cognitif a-t-il des concepts et ensuite des états mentaux d'ordre supérieur, ou bien possède-t-il directement des états mentaux d'ordres supérieurs qui sont composés en partie de concept ? Y a-t-il un concept sans croyance ?

L'étude présentée ici reste ambiguë sur cette question. L'analyse abstraite du problème et la structure proposée du concept, sont théoriques et *a priori*, ce qui permet de parler du concept de manière isolée. D'un autre côté, la théorie *pragmatique* mise en avant ici, en insistant sur la nature inférentielle du concept, suppose que le concept est toujours utilisé avec d'autres concepts, c'est-à-dire au niveau des conceptions, et si les états mentaux se situent au niveau des conceptions, cela signifie que le concept est toujours utilisé, pratiquement, dans une croyance, un jugement, un désir, un plan d'action ou quelque chose de cet ordre-là. C'est la position que je tiendrais pour le moment sur cet aspect de la question.

Reste aussi celle de savoir si les agents cognitifs dont il est constamment dans cette étude sont des agents humains ou non. Évidemment ce sont des agents humains, du moins je ferai l'hypothèse que les humains possèdent des concepts. Mais les concepts AGENT COGNITIF et AGENT HUMAIN ne sont pas strictement synonymes, et rien n'empêche que d'autres espèces – en particulier parmi les grands primates – soient composées d'agents cognitifs, du

moment que ces agents soient capables de répondre aux réquisits de possession du concept, en particulier avoir une théorie naïve de l'esprit qui leur permet d'attribuer à autrui des représentations mentales similaires aux leurs. Cette remarque tient compte de la distinction que je fais entre les contenus conceptuels et non-conceptuels, et en particulier avec la notion de *percept* qui est non-conceptuelle mais permet déjà une forme de catégorisation et d'inférence à un niveau différent que le niveau conceptuel. D'autres espèces, et en principe toute entité dotée d'un système d'échange d'information avec son environnement – que ce soit une anémone ou un robot – possède un traitement non-conceptuel du contenu de cette information, et si elle est dotée d'un système de mémoire et de récupération de l'information stockée dans cette mémoire, elle peut même posséder des percepts. La possession de concept requiert un traitement de la représentation au sens de correction d'évaluation d'application ou d'utilisation du traitement de l'information. Cette conception du concept disqualifie un certain nombre d'entités comme étant pourvues de concepts. Par exemple une machine capable de classer automatiquement des échantillons sonores suivant leurs morphologies ne possède pas nécessairement de concepts, mais est un système de catégorisation automatique performant, c'est tout. Cela ne veut pas dire que l'ours polaire qui me regarde ne possède pas le concept BRUYANT AVEC SON FUSIL ou BON À MANGER. Le fait est que je reste neutre sur ce point dans ce travail, et que cette neutralité ne nuit, ce me semble, à aucun moment, à l'analyse théorique et normative du concept, telle qu'elle est menée ici.

Chapitre 2

Analyse du problème

2.1 Contenu non-conceptuel et contenu conceptuel

Dans une expérience, un agent cognitif traite deux types de contenus différents au moins. Le premier sera dit être *non-conceptuel* et le second *conceptuel*.

Soit l'expérience suivante :



FIG. 2.1 – Illusion de constance de taille (image : David Falco)

Deux personnages se tiennent dans un paysage. Les deux personnages sont perçus comme ayant des tailles différentes.

Le contenu non-conceptuel dans cette expérience correspond aux données de la perception qui permettent à l'agent cognitif de voir quelque chose dans l'image. L'agent voit deux personnages du fait d'un certain agencement de formes et de contrastes qui sont captés et traités par son système visuel.

Le contenu conceptuel de cette expérience correspond aux jugements, croyances, désirs, actions ou tout autre état mental que l'agent peut entretenir du fait cette expérience. Par exemple l'agent peut dire qu'il y a deux personnages dans l'image, que le paysage derrière est celui du nord de la Norvège ou bien que les deux personnages sont de tailles différentes.

Imaginons qu'un agent, face à cette image, pense que les deux personnages sont de tailles différentes.

Imaginons maintenant que des informations complémentaires sont données à propos de cette image. Une légende pourrait par exemple indiquer qu'il s'agit d'une illusion de perception, ou bien quelqu'un pourrait dire à l'agent qu'en fait les deux personnages dans l'image ont la même taille, ou bien le fond de l'image pourrait disparaître et ne laisser visible que les deux personnages :



FIG. 2.2 – Constance de taille (image : David Falco)

Ces nouvelles informations peuvent faire changer le jugement que l'agent porte sur l'image. À partir du contraste entre les deux types d'information, l'agent peut revenir sur son jugement selon lequel les deux personnages sont de tailles différentes, et dire maintenant qu'ils sont de la même taille.



FIG. 2.3 – Contraste dans l'illusion de constance de taille (image : David Falco)

En y regardant de plus près, on s'aperçoit que l'illusion de constance de taille perdure malgré ces nouvelles informations. Le jugement de l'agent a changé, mais cela n'a pas affecté l'impression laissée par l'expérience perceptive.

Cette expérience montre la distinction entre le contenu non-conceptuel, qui se situe au niveau des informations captées et traitées par le système perceptif de l'agent, et le contenu conceptuel qui se situe au niveau de la pensée et des états mentaux entretenus par cet agent à propos de cette expérience.

2.2 Différence entre le non-conceptuel et le conceptuel

2.2.1 Représentation

La principale différence entre le contenu non-conceptuel et le contenu conceptuel est que le second est *représentationnel* alors que le premier ne l'est pas.

La *représentation* est le fait de rendre un élément présent à quelqu'un en le montrant ou en le faisant savoir, à travers la restitution de traits fondamentaux de cet élément. La représentation est donc une fonction qui assigne un code à un élément. Ce code se base sur les traits fondamentaux qui caractérisent cet élément. De fait, α peut représenter l'élément a si α est caractérisé par la propriété p qu'instancie a . α sera dit représenter a s'il existe une fonction telle que :

$$f : \alpha \mapsto a : \begin{cases} 1 & \text{si et seulement si } a \text{ instancie } p \\ 0 & \text{sinon} \end{cases} \quad (2.1)$$

Le contenu représentationnel est donc descriptible en termes propositionnels Peacocke [1983, p. 8], et l'agent ne peut posséder ce contenu que s'il est en mesure de le décrire, ce qui satisfait à la contrainte de possession d'un concept, telle qu'elle est donnée par Peacocke [1992]. Cette contrainte affirme qu'un agent possède un concept s'il possède un état cognitif qu'il ne peut posséder sans pouvoir être en mesure de le décrire.

Mais un caractère plus intéressant encore découle de la formulation de la représentation telle qu'elle est énoncée ici. Si la représentation est une fonction qui attribue à un élément un code en fonction des caractères fondamentaux que possède cet élément – par exemple le concept CHEVAL représentera Rossinante du fait que Rossinante possède les caractères de la équidité – cela signifie que la fonction de représentation est elle-même évaluable. C'est-à-dire qu'il est possible de savoir si α représente bien a .

L'évaluation de la fonction d'assignation, est appelée « correction ».

La correction est l'action de rectifier, d'amender ou de ramener à une règle ou une norme un élément s'il ne s'y conforme pas.

L'évaluation de la fonction d'assignation f se fait sur la base du caractère p que a doit vérifier pour être α . Pour toute fonction de représentation f il existe donc une *fonction de correction* ou d'évaluation \mathfrak{f} qui permet de dire si f s'applique correctement entre a et α .

$$\mathfrak{f} : f \rightarrow \begin{cases} 1 & \text{si et seulement si } a \text{ instancie } p \\ 0 & \text{sinon} \end{cases} \quad (2.2)$$

C'est du fait de cette fonction de correction, qui caractérise la représentation, que la représentation peut-être corrigée et modifiée en fonction des nouvelles informations et la mise en évidence d'une erreur.

En comparant les deux images, celle des deux personnages avec le fond et celle des deux mêmes personnages sans le fond, l'agent peut se rendre compte qu'en fait les deux personnages ont la même taille, mais qu'une illusion intervient lorsque le fond est ajouté. Cette illusion est causée par le conflit entre les propriétés saillantes, qui correspondent aux traits caractéristiques qui interviennent dans la correction, propriétés saillantes des deux personnages d'une part, et celles de la profondeur issue des lignes fuyantes d'autre part. L'illusion disparaît lorsque le conflit est annulé, c'est-à-dire lorsque le fond est supprimé. Cependant elle réapparaît et persiste lorsque le fond est remplacé. Le jugement a été modifié entre-temps, c'est-à-dire la représentation que l'agent a de la scène, mais pas l'impression par la perception. De fait le contenu conceptuel peut-être corrigé, parce qu'il est représentationnel. Le contenu non-conceptuel lui, n'est pas modifié sans une intervention sur l'expérience elle-même – en supprimant le fond par exemple – et donc n'est pas représentationnel.

Si les états mentaux d'ordre supérieurs, tels que les jugements, croyances, connaissances, désirs, plans d'action ou autres peuvent être modifiés et corrigés, c'est parce qu'ils sont composés de concepts. Le concept, contrairement aux éléments non-conceptuels, comme le percept par exemple, est donc une entité représentationnelle. C'est parce qu'il peut être corrigé et modifié qu'il est possible de parler de changement de concept.

La condition de correction du concept ne conduit cependant pas nécessairement à endosser une théorie vérificationniste au sens fort du terme. Une théorie vérificationniste consisterait à dire que pour posséder un concept, l'agent devrait savoir sous quelles conditions quel concept s'applique à tel objet. Or cela signifierait que l'agent aurait un accès à la propriété de l'objet indépendamment du concept, et en ce sens le concept deviendrait inutile. Cela serait le cas si l'agent était omniscient ou assuré de la vérité de sa connaissance. Le fait qu'un agent puisse se tromper, en pensant que les baleines sont des poissons ou en prenant un chien pour un loup, montre que l'erreur est possible.

Une théorie vérificationniste du concept s'oppose à une théorie relativiste qui vérifierait la *thèse de la ratification* de Peacocke [1992] :

Ratification Thesis : When someone possesses a particular concept, the correctness of his judgement of whether a given object falls under the concept is not determined by any of the circumstances surrounding his judgement of whether others fall under it ¹. [Peacocke, 1992, p. 193]

C'est-à-dire que la vérité de l'assignation du concept ne reposerait pas sur les propriétés des objets, mais sur la simple utilisation du concept par l'agent. Cette thèse refuse une objectivité de la signification et du contenu du concept. Or si les baleines sont des mammifères, c'est certainement bien du fait d'une certaine configuration de l'état du monde dans lequel ce concept est utilisé.

Une théorie qui respecterait la thèse de la ratification s'appuie la vérité de l'application du concept sur les autres concepts et état mentaux possédés par l'agent. C'est-à-dire que l'agent pense que les baleines sont des poissons du fait ce qu'il pense des poissons et ainsi de suite. Cela conduit à un holisme général, et tombe sous le joug des critiques faites par Fodor and Lepore [1992] selon lequel ce type de théorie ne décrit d'autre que ce qu'elle veut bien décrire. Le problème est que si les baleines sont des mammifères, la théorie peut bien dire que ce sont des poissons, il n'en reste pas moins qu'elle dira quelque chose de faux. Le vérificationnisme entendu sous cette forme est en fait l'importation de la théorie de la preuve mathématique, dans la théorie des concepts. Dans une théorie mathématique de la preuve, un énoncé est vrai s'il découle logiquement d'autres énoncés considérés comme axiomes. Une théorie des concepts calquées sur la théorie de la preuve ou une théorie axiomatique, est bien trop naïve, car elle ne tient ni compte de l'état du monde, ni compte de l'état de l'agent. En somme cela revient à dire qu'un agent est une machine qui se contente de traiter des énoncés. Une théorie de l'esprit basée sur cette conception peut se trouver chez Turing [1950]. Cependant l'esprit n'est pas une machine. Il doit prendre en compte l'environnement et le rapport que l'agent a avec lui. De fait une théorie

¹ «**Thèse de ratification** : Lorsque quelqu'un possède un concept particulier, la correction de son jugement selon lequel un objet donné tombe ou non sous le concept n'est déterminée par aucune des circonstances entourant son jugement selon lequel d'autres objets tombent ou non sous ce concept.» [NdA toutes les traductions des citations sont personnelles, l'original faisant référence.]

des concepts ne peut pas accepter telle quelle la thèse de la ratification [Peacocke, 1992, p. 194].

Cela n'empêche pas que le concept ait des conditions d'application et que ces conditions doivent se conformer à un principe de correction. Ce que cela signifie c'est simplement que l'intension du concept n'est pas suffisante pour déterminer le concept. Cependant la référence seule n'est pas non plus suffisante puisque l'agent n'est pas omniscient. D'un côté il y a l'écueil du holisme, de l'autre celui de l'omniscience.

La seule chose que nous pouvons dire, c'est que le concept se tient entre ces deux extrêmes. Quoi qu'il en soit, la représentation n'est pas suffisante pour comprendre le concept.

2.2.2 Caractère épistémique

Si le concept est représentationnel au sens donné précédemment (2.2.1), cela implique que l'agent doit avoir accès au concept qu'il possède. Cet accès est le caractère épistémique du concept.

Le caractère épistémique a deux formes : une faible et une forte.

2.2.2.1 Caractère épistémique fort

Un état mental est épistémiquement fort si l'agent qui le possède a accès à ce contenu et possède la connaissance de cet accès, c'est-à-dire qu'il peut donner les raisons de cet accès et de cette possession ².

2.2.2.2 Caractère épistémique faible

Un état mental est épistémiquement faible si l'agent qui le possède a accès à ce contenu mais ne possède pas la connaissance de cet accès, c'est-à-dire qu'il ne peut pas donner les raisons de cet accès ³.

Le caractère épistémique fort correspond à la définition classique de la connaissance comme croyance vraie justifiée. Le caractère épistémique faible correspond à une connaissance au sens commun du terme, au sens d'avoir accès à quelque chose ou être conscient de cette chose sans être en mesure de le justifier. Cette forme de connaissance est celle que manifeste un agent

²Cela correspond à un critère *internaliste* de la connaissance.

³Cela correspond à un critère *externaliste* de la connaissance.

lorsqu'il dit que ce qui se tient devant lui est une chaise, sans qu'il soit capable de donner la définition exacte (si tant est qu'il y en ait une) de ce qu'est une chaise.

La distinction entre les deux niveaux d'épistémicité recoupe la distinction faite par Dretske [1969] entre épistémique et non-épistémique. Cette distinction correspond à la différence qu'il y a entre le fait de voir quelque chose et le fait de voir *que* c'est quelque chose. Cependant, la distinction telle qu'elle est proposée par Dretske n'est pas suffisante. Elle ne permet pas de rendre compte de la différence qu'il existe entre le fait d'être en mesure de dire pourquoi ce que l'on voit comme rouge est rouge. C'est-à-dire de rendre compte du fait d'une connaissance non explicitement justifiée. Le type d'épistémicité faible présentée ici est sans doute trop faible pour formuler la connaissance au sens classique du terme. Il permet cependant d'expliquer en quoi une croyance est déjà une prise de position sur le monde et un découpage de l'environnement qui peut servir de base à la connaissance, c'est-à-dire au sens fort de l'épistémique donné ici.

Le concept, en tant qu'il est représentationnel doit satisfaire le caractère épistémique faible au moins. Demander qu'il satisfasse également le caractère épistémique fort serait trop exigeant, car cela reviendrait à considérer que le concept est déjà une connaissance, or tous les états mentaux ne sont pas des connaissances, et si le concept est le constituant de ces états mentaux, il ne peut être si fort épistémiquement.

Si le concept est épistémique, et si cette épistémicité faible est la marque de la représentation, cela signifie que le contenu non-conceptuel n'est pas épistémique. Mais il y a dans l'expérience une attention au niveau non-conceptuel qui si elle n'est pas épistémique est cependant accessible à l'agent. Dans l'image des deux personnages avec le fond (figure 2.1), la perception discrimine bien les personnages du fond sans nécessairement les représenter comme étant des personnages.

Deux niveaux de caractères non-épistémiques doivent donc être distingués.

2.2.2.3 Caractère non-épistémique fort

Un état mental est non-épistémiquement fort si l'agent qui le possède n'a pas accès à ce contenu mais ne possède aucune la connaissance de cet accès, c'est-à-dire qu'il ne peut en donner les raisons.

2.2.2.4 Caractère non-épistémique faible

Un état mental est non-épistémiquement fait si l'agent qui le possède peut faire attention à ce contenu mais ne possède aucune la connaissance de cet accès, c'est-à-dire qu'il ne peut en donner, ni ne peut l'identifier.

Le caractère non-épistémique au sens fort correspond à un processus infra-personnel inconscient. Par exemple la posture est engendrée par un système vestibulo-visuo-kinestésique complexe dont l'agent n'a aucune idée lorsqu'il est en œuvre. Dans l'expérience de l'image avec les deux personnages, cela correspond aux données physiques captées et traitées par le système visuel de l'agent. L'agent ne peut en aucun cas modifier ou changer ces informations sans devoir agir sur le mode même de présentation, c'est-à-dire sur le contexte de l'expérience même. Il est certain que si l'agent éteint la lumière, son expérience de l'image sera différente, mais cela correspond à une autre expérience et non pas à un changement de position de sa part quant à l'image.

Le caractère non-épistémique au sens faible est plus compliqué à saisir. Sa formulation même le rapproche de la formulation du caractère épistémique faible et il n'est pas évident que l'on puisse distinguer les deux. Cependant la différence entre les caractères faibles de l'épistémique et du non-épistémique se comprend mieux si on les éclaire avec la notion de représentation donnée précédemment. Il y a dans l'expérience des données dont l'agent qui les possède est conscient sans qu'il puisse agir dessus. Dans un raccourci rapide on dira que l'agent est passif quant au non-épistémique alors qu'il est actif dans quant à l'épistémique. Un agent peut être attentif au fait qu'un arbre soit vert par exemple, c'est-à-dire de discriminé la couleur verte de l'arbre de la couleur bleu du ciel sans pour autant savoir que c'est du vert ou de quelle nuance est cette couleur. La discrimination est donc plus faible que l'identification telle qu'elle est exigée par la représentation et le caractère épistémique. Il faut d'ailleurs noter que c'est du fait de cette identification, et donc de la représentation et du caractère épistémique, que la contrainte de généralité de [Evans \[1982\]](#) s'applique au concept.

Thus, if a subject can be credited with the thought that a is F , then he must have the conceptual resources for entertaining the thought that a is G , for every property of being G of which he has a conception. This

is the condition I call ‘The Generality Constraint’⁴. [Evans, 1982, p. 104]

Il y a une différence entre voir du vert et voir que c’est du vert sans être capable de dire pourquoi c’est du vert. C’est cette différence qu’il y a entre le caractère épistémique faible et le caractère non-épistémique faible.

2.2.3 Sens/Référence

Le concept est donc une entité représentationnelle et épistémique au moins au sens faible du terme. Le concept CHEVAL s’applique à Rossinante du fait qu’il représente Rossinante et que l’agent qui le possède a accès à cette représentation. Mais ce n’est pas suffisant pour caractériser le concept. Le concept CHEVAL s’applique bien à Rossinante si Rossinante est un cheval, c’est-à-dire qu’il manifeste toutes les qu’un objet du monde doit manifester pour être un cheval. Le concept CHEVAL est donc une fonction qui s’applique aux objets de l’environnement qui sont des chevaux :

$$\text{CHEVAL} \mapsto x \begin{cases} 1 & \text{si } x \text{ est un cheval} \\ 0 & \text{sinon} \end{cases} \quad (2.3)$$

Cette assignation correspond à la *dénotation* du concept, c’est-à-dire à sa *référence*. En somme la référence du concept est l’ensemble de tous les éléments dans l’environnement dans lequel est appliqué ce concept qui rend vrai la propriété qui est associée à ce concept. La référence du concept peut se ramener à l’extension du prédicat de cette propriété. Par exemple, les chevaux sont tous les éléments du monde qui vérifient le prédicat « être un cheval ».

Mais la dénotation dépend de l’état actuel de l’environnement auquel le concept est appliqué. Il se pourrait très bien que Rossinante ne soit pas en fait un cheval, et alors le concept cheval ne s’y appliquerait pas. Ou bien si Rossinante est un cheval et donc que c’est un mammifère parce que les chevaux sont des mammifères, cela est vrai du fait de l’agencement des êtres vivants dans cette configuration-là du monde, mais il pourrait très bien se faire que les chevaux soient des poissons, et alors dire que Rossinante est

⁴“Ainsi, si un sujet peut être crédité de la pensée que a est F , alors il doit posséder les ressources conceptuelles pour recevoir la pensée que a est G , pour toute propriété d’être G de ce dont il a la conception. C’est cette condition que j’appelle ‘la contrainte de généralité’.”

un mammifère parce que c'est un cheval serait faux. Il faut donc distinguer l'extension, ce que le concept dénote dans le monde, de la manière dont il le fait, c'est-à-dire de son sens ou de son intension.

L'intension n'est pas la référence actuelle du concept, mais l'ensemble des conditions qui permettent au concept de s'appliquer à l'objet, c'est-à-dire les possibilités de vérité du concept [Gallin, 1975]. Le concept CHEVAL s'applique aux éléments qui ont telles et telles propriétés qui font d'eux des chevaux. De cette manière on ne parle pas d'objet, mais de propriétés, de traits fondamentaux qu'ils doivent satisfaire pour tomber sous le concept. La liste de ces propriétés est la *fonction caractéristique* du concept. Et cette fonction caractéristique correspond au terme abstrait associé au concept [Church, 1956, Prior, 2002]. Terme abstrait car il ne dépend d'aucun élément présent dans l'environnement, mais permet de caractériser des éléments de cet environnement. Le terme abstrait 'cheval' est ce qui peut se penser de Rossinante et de Joly Jumper, sans qu'il ne soit spécifiquement l'un ou l'autre. La fonction caractéristique est notée :

$$\lambda\Phi \tag{2.4}$$

où λ est l'opérateur d'abstraction, et Φ la liste des propriétés que doit satisfaire un élément pour être caractérisé par cette fonction.

2.3 Concept

Le concept doit se conformer aux réquisits de la représentation, de l'épistémicité et de la distinction entre le sens et le référence. Cependant, ces caractères ne sont pas suffisants pour rendre compte du concept. En effet, si le concept est le composant des états mentaux tels que les croyances, connaissances, désirs, ou plans d'action, alors il faut que la structure du concept lui permette d'entrer dans la composition de ces états mentaux.

2.3.1 Individuation du concept

Si un concept entre dans la composition d'un état mental alors il doit pouvoir être individué. Dans le vocabulaire de Fodor [1998], dire qu'un concept est indivudable c'est dire que c'est un particulier mental :

Concepts are mental particulars ; specially, they satisfy whatever ontological conditions have to be met by things that function as mental causes and effect ⁵. [Fodor, 1998, p. 23]

Fodor considère, de par son nativisme que l'individuation du concept est naturelle, du fait de la relation causale entre le concept en tant qu'entité mentale et l'objet dans le monde auquel il réfère. L'agent cognitif qui possède le concept ne joue aucun rôle dans l'individuation de celui-ci. Cependant il est difficile de voir quelle est l'individuation du concept BALEINE dans l'esprit d'un agent qui pense que ce sont des poissons. Est-ce qu'il possède le concept BALEINE en tant qu'il désigne des animaux dans le monde qui sont des mammifères et sans trop savoir comment il se fait que dans l'esprit de cet agent, ce concept désigne des poissons, ou bien est-ce que cet agent ne possède pas le concept BALEINE mais le concept PALEINE qui désigne des animaux qui ont toutes les propriétés des baleines excepté le fait que ce sont des poissons, ou bien est-ce que l'agent possède le concept BALEINE qui dénote bien des animaux qui sont des mammifères, mais l'agent individue mal ce concept en pensant que ce sont des poissons ?

Selon Peacocke [1992], un concept est individué par ses conditions de possession.

Simple Account : When a thinker possesses a particular concept, an adequate psychology should explain why the thinker meets the concepts possession condition “**Compte simple** : Quand un penseur possède un concept particulier, une psychologie adéquate devrait expliquer pourquoi le penseur atteint les conditions de possession des concepts.”. [Peacocke, 1992, p. 177].

Les conditions de possession du concept indiquent l'application du concept – quand celui-ci s'applique à l'objet – et son usage – ses relations aux autres concepts [Peacocke, 1992, p. 182]. De fait, un principe d'individuation du concept peut se faire à partir des conditions d'application et d'usage du concept.

Concept F is that unique concept C to possess which a thinker must meet condition $\mathcal{A}(C)$.⁶. [Peacocke, 1992, p. 6].

⁵“Les concepts sont des particuliers mentaux ; plus spécialement, ils satisfont les conditions ontologiques, quelles qu'elles soient, qui doivent être remplies par les choses qui fonctionnent comme cause et effet mental.”

⁶“Le concept F est cet unique concept C que pour posséder un agent doit rencontrer le condition $\mathcal{A}(C)$.”

où C est une variable sur les concept et \mathcal{A} une lettre schématique.

Chaque condition spécifique d’usage et d’assignation détermine donc un unique concept. Cette individuation du concept est différente de celle proposée par Fodor [1998], puisque les conditions d’individuation ne se fondent pas uniquement sur la référence du concept, mais également sur la prise en compte des états mentaux de l’agent qui possède ce concept. La possession d’un concept peut donc modifier l’application ou l’usage d’un autre concept. C’est parce que l’agent pense que les poissons sont tout les animaux qui vivent dans l’eau et qui n’ont ni ailes ni pattes, qu’il pense que les baleines sont des poissons. C’est-à-dire que son application et son usage du concept POISSON a des répercussions sur l’usage et l’application de son concept BALEINE. Donc les concepts ne sont pas indépendants les uns des autres, comme le soutient l’atomisme de Fodor, mais sont reliés, au moins partiellement, entre eux, comme l’affirme le holisme local de Peacocke.

Formulation for Local Holism : Concepts F_1, \dots, F_n are those concepts C_1, \dots, C_n to possess which a thinker must meet the conditions $\mathcal{A}(C_1, \dots, C_n)$ ⁷. [Peacocke, 1992, p. 10]

Ces relations entre les concepts indiquent que la possession d’un concept peut modifier la possession d’un autre concept. Ce holisme local, c’est-à-dire que plusieurs réseaux de concepts peuvent coexister dans une conception qui elle doit être consistante, permet d’expliquer l’existence de domaines à l’intérieur d’une conception. Ces domaines correspondent aux théories naïves du sens commun. Dans une biologie naïve par exemple, tout animal qui vit dans l’eau, qui nage et qui n’a ni aile ni patte, est un poisson, et en ce sens les baleines sont des poissons. Ce n’est que lorsque l’on prête attention à des caractères qui ne sautent pas nécessairement aux yeux, que l’on développe une théorie scientifique.

Le concept dépend d’une part de son application, c’est-à-dire de sa référence, et d’autre part de son usage, c’est-à-dire des relations qu’il entretient avec les autres concepts. Pour rendre compte du concept il est donc nécessaire d’expliquer ces deux aspects du concept.

⁷“**Formulation pour le holisme partiel** : Les concepts F_1, \dots, F_n sont ces concepts C_1, \dots, C_n que pour posséder l’agent doit rencontrer les conditions $\mathcal{A}(C_1, \dots, C_n)$ ”.

2.3.2 Catégoricité du concept

La référence du concept est souvent considérée, dans la littérature, comme une forme de catégorisation. Dire que les éléments a_1, \dots, a_n relèvent du concept α est pensé d'un point de vue ensembliste : les éléments a_1, \dots, a_n sont inclus dans l'ensemble dont le label est α .

Dans la mesure où le concept a une extension, l'interprétation ensembliste va de soit. Cette interprétation est toute fois étendue en un sens plus large. Le concept est une catégorie en tant qu'il est un découpage de l'environnement qui permet à l'agent qui le possède de comprendre cet environnement. Par exemple, un rapide coup d'œil dans la pièce me fait voir qu'il y a des livres, des étagères, une table et une chaise. Pour comprendre cette scène, il me faut posséder les concepts LIVRE, ÉTAGÈRE, TABLE et CHAISE, et que les données issues de ma perception au niveau non-conceptuel s'arrangent pour activer ces concepts. Je perçois une certaine forme, et cette forme correspond à un élément subsumé par le concept LIVRE.

Le caractère catégoriel du concept concerne son extension mais également sa fonction cognitive : le rôle du concept est de permettre de composer des états mentaux d'ordre supérieur comme les croyances, les connaissances, les désirs ou les plans d'action. Le concept est le découpage du monde en éléments sur lesquels ces états mentaux vont s'attacher. Cet aspect du concept étant jugé central dans la littérature sur le domaine, une étude du concept se transforme donc assez rapidement en une étude sur la classification. Un exemple frappant de ce type est les tâches demandées dans les expériences psychologiques. Le plus souvent il s'agit de demander si l'objet x présenté est un α ou un β . Tâche qui se résume à mettre l'élément en question dans l'ensemble qui lui correspond.

Cependant, une étude rigoureuse de la classification montre les problèmes formels qui y sont liés [Gordon, 1981, Estes, 1994]. Le problème le plus embêtant est qu'aucune classification ne peut être exhaustive. Ce problème est en fait une forme du problème de l'auto-référentialité, plus connu sous le nom de *paradoxe de Russell*. Ce paradoxe dit simplement que si tous les objets du monde peuvent être mis dans des ensembles ou des classes, la classe qui contient tous ces objets est un objet du monde qui n'entre dans aucun ensemble, donc tous les objets du monde ne sont pas compris dans un ensemble.

Il y a deux manières d'aborder la catégorisation. La première est dite *bottom up* (de bas en haut), la seconde, *top down* (de haut en bas). L'idée est assez simple : chaque objet appartient à un ensemble d'objets à un niveau et il est possible d'inclure des ensembles d'objets dans des ensembles à un niveau supérieur, schématiquement cela donne un graphe orienté ou bien un arbre.

Dans la méthode *bottom up*, les éléments du monde sont inclus dans l'un des ensembles en fonction de leur ressemblance. C'est-à-dire que des éléments qui se ressemblent seront regroupés ensemble pour faire une classe. La ressemblance entre les éléments est calculée à l'aide d'une fonction de similitude appelée *principe contrastif* [Tversky, 1977] :

$$(\text{Principe contrastif}) \quad Sim(I, J) = af(I \cap J) - bf(I - j) - cf(J - i) \quad (2.5)$$

où I et J sont respectivement l'ensemble des traits caractéristiques de i et de j , et a , b , et c sont les valeurs pondérées de différents traits distinctifs.

Cette méthode est utilisée par la théorie empirique des concepts défendue par Prinz [2002], et les théories des exemplaires et des prototypes.

L'inconvénient de cette approche est qu'elle maltraite les cas particuliers. Certains éléments relevant bien d'un concept ne partagent parfois que très peu de traits caractéristiques avec les autres éléments subsumés par le concept. Si l'on considère certains traits jugés caractéristiques, les baleines sont plus proches des requins ou des dorades que des chiens et des chats. Le premier problème est donc choix – et par conséquent la pertinence – des traits caractéristiques et de leur pondération. La cladistique et la systématique connaissent bien ce problème. Plusieurs cladogrammes peuvent être établis à partir d'un même ensemble de caractères. Il s'agit donc de choisir l'un de ces graphes, ce qui rend relatif la catégorisation. D'un autre côté, le choix dans les traits caractéristiques et leur pondération est évident lorsque l'on considère que la catégorisation, et donc le découpage du monde à l'aide de concepts, est faite en fonction d'un but et d'une utilité. L'intérêt de posséder le concept TABLE et de pouvoir agir d'une certaine manière avec les objets du monde qui sont des tables.

Un autre problème est l'homologie entre des traits caractéristiques d'ensembles disjoints. D'un certain point de vue il y a un point commun entre un Airbus A-380 et un moineau : les deux ont des ailes. Évidemment les ailes

d'un avion n'ont pas grand-chose à voir avec celle d'un oiseau, cependant, si le trait caractéristique retenu est « ce qui permet de planer dans l'air », alors les avions et les oiseaux se retrouveront dans une même catégorie. Et si en plus de ce caractère, celui d'être un objet manufacturé ou un être vivant est pris en compte, alors les éléments compris dans les êtres vivants auront un point commun avec les objets manufacturés, ou bien certains éléments qui sont des êtres vivants ne seront pas pris en compte parce qu'ils n'ont pas l'aile. Le problème non seulement du choix et de la pertinence des traits caractéristiques se pose, mais également la question de la cohérence des traits entre eux. La consistance de l'ensemble implique une forme de holisme.

Or pour résoudre le problème de la cohérence et de la consistance des caractères, il ne faut plus tenir compte des éléments dans le monde, mais des relations entre caractères. C'est là l'objet de la méthode *top down*. Avant de considérer les éléments à catégoriser, la méthode descendante préconise de se doter d'un ensemble consistant de caractère et de les appliquer ensuite au monde. Le problème devient alors de fait de savoir si une description ne s'applique pas à l'environnement, c'est du fait de la pauvreté de l'environnement ou bien de l'impossibilité d'aucun élément à s'y conformer quel que soit l'environnement. Par exemple qu'en est-il des dahus et des licornes ?

L'idéal sans doute serait d'associer les deux méthodes, cependant cela conduit à des paradoxes, paradoxes qui se rapprochent de celui de Russell.

Imaginons qu'il soit possible d'assigner un label unique pour chaque élément du monde que l'on cherche à catégoriser. Un label de ce type est un nombre de Gödel ou un nom propre par exemple. Autrement dit, on établit la liste exhaustive des éléments du monde à la manière d'un plus pur nominalisme. Cela signifie qu'un élément du monde est représentable si et seulement s'il existe une fonction f qui assigne à chacun des éléments de l'environnement un tel label, et que sinon cet élément n'appartient pas à cet environnement :

$$\begin{aligned} & \forall v[\varphi(v, n_1, \dots, n_k) \leftrightarrow v = f : l \mapsto (n_1, \dots, n_k)] \\ & \text{si } (n_1, \dots, n_k) \in \mathcal{E}, \text{ alors } \varphi(n_1, \dots, n_k) \\ & \text{si } (n_1, \dots, n_k) \notin \mathcal{E}, \text{ alors } \neg\varphi(n_1, \dots, n_k) \end{aligned} \quad (2.6)$$

Un élément de (n_1, \dots, n_k) appartient à l'environnement \mathcal{E} si la fonction f lui assigne un label l , sinon il n'y appartient pas. Mais s'il est possible de dire qu'un élément n'appartient pas à l'environnement \mathcal{E} c'est qu'un label lui a

attribué, et donc qu'il entre dans la description des éléments qui ne décrivent pas cet environnement. Donc la fonction f qui assigne à cet élément le label indiquant qu'aucun label ne s'applique à cet élément :

$$\forall v[\varphi(v, l) \leftrightarrow v = \neg\varphi] \quad (2.7)$$

d'où

$$\varphi = \neg\varphi \quad (2.8)$$

ce qui est contradictoire [Tarski et al., 1953].

La conclusion – qui est similaire à celle du théorème de Ramsey – indique tout simplement qu'il n'existe pas de méthode exhaustive de catégorisation. De fait un choix doit être fait au départ sur la manière dont l'environnement est conceptualisé.

Le concept cependant, contrairement à une classification scientifique ou formelle, est une entité cognitive. La fonction du concept au sein de la cognition peut déterminer la catégorisation.

Les éléments déterminés par les concepts proviennent en partie au moins du donné de l'expérience, c'est-à-dire du traitement non-conceptuel. Cela permet de mettre en évidence le caractère inné de certains concepts au moins. Le concept devient alors une catégorisation du contenu de la perception qui se déclenche automatiquement sous forme de *routines* [Fodor, 1998, p. 24].

La notion de routine, appliquée au concept pose un certain nombre de problèmes. Une routine est un processus qui s'enclenche de manière automatique dès qu'apparaissent certaines conditions stimuli. Cette mise en route basée sur un schéma cause/effet convient parfaitement à la sémantique causale du concept telle qu'elle est défendue par Fodor. Comme pour Fodor il y a un lien direct et nécessaire entre le concept et l'objet qu'il désigne, il est naturel de penser que lorsque l'objet apparaît, le concept apparaît lui aussi.

D'une certaine manière la conception de Peacocke du concept permet une interprétation de celui-ci en terme de routine. L'agent perçoit d'abord une scène qui est décrite en termes de proto-propositions puis en termes propositionnels et donc conceptuels. La différence est que le lien causal est indirect et passe par ce premier découpage, ce qui permet de penser qu'un

objet rectangulaire pourrait être décrit proto-propositionnellement comme un cube, et être interprété ensuite comme étant une table, ou bien comme étant une caisse. La relation indirecte entre l'objet de la perception et l'objet conceptualisé laisse une place à l'erreur ou au changement. L'agent peut d'abord penser que c'est une caisse puis ensuite penser que c'est en fait une table sans que l'objet de l'expérience ait changé. Si le lien causal est direct, comme dans la conception atomiste de Fodor, l'agent doit penser directement que c'est une table si l'objet en question est une table. Il serait contradictoire que deux concepts différents puissent s'appliquer causalement et directement à un même objet.

L'automaticité de la routine suggère que l'agent n'est pas conscient – au sens de porter attention – au fait qu'il catégorise l'environnement de la manière dont il le fait. En un sens cela est indéniable, lorsque je pose un gobelet sur la table je ne me dis pas en même temps « ceci est un gobelet, ceci est une table », je fais une action plus ou moins automatiquement à l'aide d'éléments que je repère et identifie de manière automatique. Ce type d'interaction avec le monde peut se faire avec des objets manufacturés complexes. Avec l'habitude, je tape sur le clavier sans nécessairement regarder les touches, ou bien je puisse conduire et changer de vitesse sans regarder le pommeau ni me dire « tiens il faudrait peut-être changer de vitesse pour ralentir ». Les réflexes conditionnés peuvent être très compliqués et se baser sur un apprentissage de normes et de règles complexes. Un escrimeur doit faire un certain nombre de gestes précis et particuliers dans un laps de temps durant lequel il ne peut pas réfléchir à son action. La routine ne signifie donc pas une forme de nativisme ou d'innéisme. Cependant, toute la question réside dans l'attention et l'accès à cette routine que peut avoir l'agent. S'il s'agit d'une routine inconsciente, alors cela signifie que l'application du concept – l'identification de l'objet perçu comme étant une table – est non-épistémique, au sens donné plus haut. Or nous avons dit que le concept devait être épistémique s'il était une représentation. Fodor [1998, p. 15] affirme que le concept est représentationnel, mais il ne reconnaît pas le caractère de correction du concept tel qu'il est rattaché dans l'interprétation donnée ici, à la représentation. Dire que le concept est une routine peut se comprendre en tant qu'il est un processus mental qui marche indépendamment de toute volonté de l'agent. De

cette manière le concept pourrait être ramené à un précâblage ou à un câblage issu de l'apprentissage. Mais alors, cela signifierait que le niveau proto-propositionnel de Peacocke correspondrait à ce niveau de routine tel que décrit par Fodor. Or pour poser un gobelet sur une table, effectivement j'ai besoin d'individuer le gobelet, la table et la relation entre les deux, mais je n'ai pas besoin de les identifier en tant que tels. C'est-à-dire que l'individuation peut-être non-épistémique et non-conceptuelle mais alors ce ne sera pas avec des gobelets ou des tables que je performerai l'action, mais avec des éléments ayant des propriétés saillantes auxquelles mon système perceptif est sensible, et qui déclencheront des routines d'action câblées ou précâblées suivant qu'elles sont innées ou apprises, à un niveau non-conceptuel. Un réflexe ne fait intervenir aucune délibération de la part de l'agent, mais un conditionnement de celui-ci, et le réflexe se met en branle si et seulement si des propriétés saillantes particulières et pertinentes sont présentes dans l'environnement compte tenu du mode de présentation de l'expérience en question. Autrement dit un réflexe peut se déclencher dans des conditions non optimales d'expérience, alors qu'il n'y a pas lieu. Mais c'est là au niveau non-conceptuel que cela se passe. Or Fodor n'accepte pas l'existence d'un tel niveau. Les routines dont il parle pourraient être en réalité non-conceptuelles, c'est-à-dire des proto-jugements [Sylvand and Gangopadhyay, submitted], et non pas des jugements.

Cependant la reconnaissance d'un objet comme étant une table ne fait pas l'objet d'une délibération. Certains traits caractéristiques des objets perçus semblent être immédiatement raccordés à un concept. Une table est immédiatement reconnue comme étant une table, donc l'objet 'table' déclenche de lui-même, de par ses propriétés saillantes, le concept TABLE. Mais si le concept est lancé sous forme de routine, il n'en reste pas moins vrai cependant qu'il peut être révisé. Quelqu'un peut poser son verre sur un objet qu'il prend pour être une table, c'est-à-dire que son concept TABLE c'est activé en présence de l'objet, et la maîtresse de maison d'intervenir en disant qu'il ne s'agit pas d'une table, mais d'une œuvre d'art contemporain, l'agent peut changer sa pensée au sujet de l'objet et donc le concept lié à celui-ci. Seule la catégorisation de l'objet par l'agent a changé, pas l'objet. Donc un même objet peut supporter deux concepts contradictoires, donc le lien référentiel entre l'objet et le concept n'est pas direct.

Si le concept est effectivement une routine, il faut distinguer les routines au niveau non-conceptuel et les routines au niveau conceptuel. Au niveau conceptuel, l'agent peut modifier la sortie de la routine en la corrigeant pour la conformer à la situation, alors qu'il ne le peut pas au niveau non-conceptuel.

2.3.3 Combinaison de concepts

Les concepts sont les constituants des croyances, des connaissances, des désirs, des plans d'action, et de tout autre état mental d'ordre supérieur. Les concepts sont donc comme des briques de pensée que l'on assemble et combine pour faire des états mentaux. Cela suppose que les concepts se composent ensemble.

La combinatoire des concepts peut se comprendre à deux niveaux : les concepts peuvent se combiner ensembles pour former des états mentaux dont la nature est différente de celle des concepts :

Compositionality : concepts are the constituents of thoughts and, in indefinitely many cases, of one another. Mental representations inherit their contents from the content of their constituents ⁸. [Fodor, 1998, p. 25].

ou bien les concepts peuvent se combiner entre eux pour former de nouveaux concepts.

La combinatoire des concepts est héritée de la conception frégréenne de la pensée selon laquelle le contenu de la pensée issu de la composition est compréhensible à partir du contenu de ses constituants. Ainsi, le terme complexe 'poisson rouge' est compris à partir des termes 'poisson' et 'rouge'. Le terme composé est dit être *complexe* ou *moléculaire* et le composant est dit *simple* ou *atomique*.

Dire que les éléments qui constituent la pensée se composent entre eux, signifie souscrire à la thèse d'un langage de la pensée [Fodor, 1974] structuré comme une grammaire générative à la Chomsky. Mais accepter cette thèse n'est pas neutre. En effet, si la pensée est structurée de manière à

⁸“Compositionnalité : les concepts sont les constituants des pensées et, dans un nombre indéfini de cas, des uns des autres. Les représentations mentales héritent leurs contenus des contenus de leurs constituants.”

permettre une combinatoire de ses composants, alors il faut accepter la structure logique qui permet cette composition. Cette structure doit être propositionnelle.

Hume avait déjà imaginé une forme de combinaison des idées en formant l'idée complexe de 'montagne d'or' à partir des idées de 'montagne' et de 'or'. Mais pour permettre une telle combinaison, il faut que les termes combinés soient complets et autonomes, c'est-à-dire des termes généraux ou, en d'autres termes, des termes abstraits. En effet, lorsque nous parlons de poissons rouges ou de montagne d'or, nous ne sommes pas en train de prendre la sole meunière de notre assiette et le rouge de la couverture du Code Civil et essayons de les coller ensemble. Ce que nous faisons c'est que nous prenons les caractères fondamentaux des poissons et ceux de la couleur rouge, c'est-à-dire les concepts correspondants à ces instances, et regardons comment ces caractères peuvent s'associer de manière consistante. Ce qui se combine donc ce n'est pas la référence des termes, mais la désignation de cette réservation, en d'autres termes la fonction caractéristique.

En effet, si les concepts complexes sont composés à partir de concepts simples, et que le concept complexe reçoit son contenu à partir de ses constituants, alors cela signifie que chaque occurrence d'un même concept est substituable par une autre dans tous les contextes où elle apparaît. De sorte que la compositionnalité des concepts doit pouvoir s'énoncer en terme de logique combinatoire, suivant les règles suivantes :

$$\begin{aligned} S &= \lambda x.(\lambda y.(\lambda z.xz(yz))) \\ K &= \lambda x.\lambda y.x \\ I &= \lambda x.x \end{aligned} \tag{2.9}$$

où S correspond à la substitution d'un terme appliqué à un autre, par un troisième ; I correspond à l'identité, c'est-à-dire qu'un terme qui substitué à lui-même redonne ce même terme ; K correspond à la concaténation : à partir de deux terme, un seul est retenu.

Fodor présente la compositionnalité du concept en termes atomistiques. Suivant cette approche les concepts sont pris de manière globale, avec leur contenu, est sont assemblés tels quels et doivent ainsi se retrouver dans le composé. Cette conception, qu'il appelle « *cartésienne* » est selon lui la seule capable d'expliquer la compositionnalité, contrairement aux vues qu'il

appelle « *pragmatiques* », conception qui prend en compte ma manière dont la référence concept est dérivée [Jacob, 2005]. C'est sur cette base que Fodor critique toutes les conceptions extensionnelles du concept, c'est-à-dire les théories qui constituent le concept à partir de ses instances notamment à travers l'expérience et la saisie qu'en a l'agent. C'est ainsi que selon lui la théorie des prototypes est incapable de rendre compte d'un concept complexe tel que POISSON ROUGE (l'exemple canonique de Fodor étant '*pet fish*' qui n'a pas d'équivalent direct en français, mais correspondrait à 'poisson de compagnie', qui typiquement est un poisson rouge).

Mais les théories pragmatiques ne sont pas aussi naïves et caricaturales que Fodor veut bien le laisser entendre. La théorie des prototypes soutient que le concept est construit par l'agent qui le possède à partir des instances de ce concept que cet agent a déjà perçues, et qu'il correspond à l'ensemble des points communs, considérés alors comme traits caractéristiques, partagés par ces instances. De sorte qu'un élément apparaît qui se tient au point d'équilibre des traits caractéristiques de ces éléments. Imaginons qu'un agent ait surtout vu des truites dans la vie et quelques poissons de mer, alors son concept POISSON aura surtout les traits des truites, et un peu des autres poissons qu'il aura vus, et si son concept ROUGE est issu surtout des tomates et des coquelicots, alors il devrait s'en suivre, que le concept complexe POISSON ROUGE devrait correspondre pour cet agent à une truite peinte de couleur rouge tomate, ce qui n'est pas vraiment ce à quoi ressemble un poisson rouge. Mais lorsque la théorie des prototypes aborde la question de la compositionnalité des concepts, elle ne dit pas qu'il faut composer les prototypes en tant qu'ils seraient des instances réifiées – d'ailleurs le prototype n'est pas nécessairement instancié, il a le statut d'abstrait et n'a de réalité que dans l'espace cognitif de l'agent qui le possède – mais que cette composition fonctionne comme l'*union* de deux prototypes, à la manière d'une intersection ensembliste. L'intersection de deux ensembles doit respecter la contrainte que l'ensemble résultant de cette intersection doit lui-même être un ensemble, en particulier être consistant. De fait ne seront combinés ensemble que les caractères compatibles de ces ensembles [Krivine, 1998]. De même, nous dit Hampton [2004], seuls les caractères pertinents et compatibles seront combinés dans la composition des prototypes. Ainsi donc, la composition de deux prototypes se fera comme la construction d'un prototype à partir des caractères pondérés

deux prototypes à combiner, ce qui donnera un nouveau prototype, avec une liste de caractères propres, qui ensuite sera appliqué au monde et corrigé suivant les occurrences rencontrées. De cette manière, le prototype « poisson rouge » réfère bien aux poissons rouges tels qu'ils tournent dans les bocaux et non pas à des truites barbouillées au jus de tomate. La critique de Fodor du traitement de la composition des concepts par les théories pragmatiques semble oublier que la constitution du concept résulte d'un processus qui est à l'œuvre même dans le cas des concepts complexes. Ce n'est pas la référence elle-même qui est composée, mais la désignation, la manière de référer. De cette manière, le contenu du concept composé ne dépend qu'indirectement du contenu de ses constituants, tout en gardant les caractères de désignation de ceux-ci. Le concept composé n'est pas alors un concept de seconde zone, mais un concept à part entière, avec ses propres règles d'application et d'usage.

Il est vrai qu'une théorie purement extensionnelle du concept – qui ne se baserait que sur instances du concept aurait du mal à rendre compte de la compositionnalité du concept. Cependant, et nous le verrons lorsque nous examinerons la théorie empirique des concepts, qu'une telle théorie aurait tout simplement du mal à rendre conte du concept tout court.

La composition passe par la fonction caractéristique du concept, et associe les caractères pertinents et compatibles de concepts pour former un nouveau concept. Imaginons que $\lambda\Phi$ et $\lambda\Psi$ soient les fonctions caractéristiques respectivement des concepts α et β , et que c_i désigne un caractère. Imaginons maintenant que :

$$\Phi := \{c_1, c_2, c_3\}$$

$$\Psi := \{c_4, c_5, c_6\}$$

$$(c_3 \wedge c_4) \rightarrow \perp$$

alors il est possible de construire une fonction caractéristique $\lambda\Omega$ telle que :

$$\Omega := \Phi \cap \Psi$$

$$\Omega := \{c_1, c_2, c_5, c_6\}$$

Cette construction est possible, et comme à tout concept correspond une fonction caractéristique (cf chapitre 4), cette fonction détermine un concept. Mais elle permet de déterminer l'usage du concept, c'est-à-dire son rôle

inférentiel, elle ne garantit que ce concept a une extension dans l'environnement de l'agent. C'est ici qu'intervient la *référence postulée*, introduite au chapitre (4), c'est-à-dire que si la fonction caractéristique est consistante, l'agent est en droit de penser qu'il est possible que dans un monde possible au moins, un objet est subsumé par ce concept. C'est le cas du concept DAHU par exemple.

Contrairement à la théorie frégréenne de la composition du contenu, ici la composition ne se fait qu'au niveau de la fonction caractéristique. Admettons que γ soit le concept composé des concepts α et β . Alors la fonction caractéristique $\lambda\Phi$ de α , tout comme la fonction caractéristique $\lambda\Psi$ de β apparaissent dans la fonction caractéristique $\lambda\Omega$ de γ , de sorte que $\lambda\Omega$ est une fonction de fonction, ou plus exactement une fonction de la valeur d'une fonction, à savoir celle $\lambda\Phi$ et de $\lambda\Psi$. $\lambda\Omega$ est elle-même la valeur de la fonction F de Φ et Ψ :

$$\text{(Fonction de composition)} \quad F\varphi\psi\chi = \varphi(\psi\chi). \quad (2.10)$$

F est dite être la *fonction de composition* [Schönfinkel, 1976].

Appliqué à l'exemple précédent, cela donne :

$$\begin{aligned} F &:= \Omega(\Phi\Psi) \\ &\lambda\Omega(\lambda\Phi\lambda\Psi) \\ \lambda\Omega &:= \lambda(\Phi\Psi) \\ \lambda\Omega &:= \lambda\{c_1, c_2, c_5, c_6\} \end{aligned}$$

Pour composer deux concepts pour en obtenir un troisième, il suffit de combiner les fonctions caractéristiques de ces concepts, en appliquant le principe de consistance. De fait deux concepts contradictoires, ou de types différents ne pourront donner une fonction caractéristique consistante et donc ne se composeront pas.

Contrairement à ce que soutient Fodor, la distinction entre un point de vue cartésien ou pragmatique n'intervient pas dans la question [Jacob, 2005], une distinction entre l'intension et l'extension serait plus appropriée ici. Seules les intensions se combinent, et non l'extension. Toutefois l'extension du concept ainsi formée appartiendra aux deux concepts combinés : un poisson rouge est bien un poisson et est bien rouge.

2.3.4 Apprentissage du concept

La combinaison de concept est importante pour expliquer l'acquisition de nouveaux concepts et de nouveaux états cognitifs d'ordres supérieurs, par l'agent.

Since our repertoire of primitive, unstructured concepts is finite, these will be at least one set of concepts that does not stand in the asymmetrical relation of dependence to any other set. It is very plausible that perceptual, relatively observable concepts are elements of one such set of concepts that does not depend asymmetrically on any others. A theory of concepts must explain the nature and the possibility of such conceptually basic sets ⁹.

[Peacocke, 1992, p.61]

Si l'agent cognitif ne possède qu'un certain nombre de concepts primitifs, c'est-à-dire qui sont donnés avec l'appareillage cognitif lui-même, cela signifie que d'autres concepts que peut posséder l'agent doivent être appris [Fodor, 1998, p. 27].

L'asymétrie de dépendance entre les concepts signifie simplement que pour posséder certains concepts, l'agent doit nécessairement en posséder d'autres – pour posséder un concept complexe, l'agent doit en posséder les composants – alors que certains concepts peuvent être possédés sans requérir la possession d'aucun autre – le cas des concepts primitifs.

Il y a deux manières de considérer les concepts *primitifs* : ou bien ils sont *innés*, ou bien ils sont *acquis* de manière directe. Le concept DUR pourrait être acquis par la simple expérience, au moins dans une acception naïve. Les concepts primitifs correspondent chez Peacocke [1992] aux concepts basés sur les proto-concepts, de découpage pré-conceptuel de l'environnement en *scènes* avec des proto-objets comme des volumes simples par exemple, qui permettent une interaction avec l'environnement sans exiger aucune connaissance de celui-ci. Ce découpage se situe à un niveau non-conceptuel de haut niveau, et fait appel à des propriétés saillantes sélectionnées par la per-

⁹“Puisque notre répertoire de concepts primitifs, non structurés, est fini, il y aura au moins un ensemble de concepts qui ne se tient pas dans la relation asymétrique de dépendance avec aucun autre ensemble. Il est très probable que les concepts perceptuels, relativement observables, soient des éléments d'un tel ensemble de concepts qui ne dépendent pas asymétriquement d'aucun autre. Une théorie des concepts doit expliquer la nature et la possibilité de tels ensembles conceptuellement basics”.

ception. Les proto-concepts sont des *percepts* [Sylvand and Gangopadhyay, submitted] (cf. Chapitre 11).

L'innéisme soutient qu'il n'est pas possible d'apprendre quelque chose à partir de rien. Or s'il faut déjà posséder quelque chose pour pouvoir apprendre autre chose, ou bien cela est possible et donc il y a des éléments que l'agent possède sans apprentissage, ou bien alors cela conduit à une régression à l'infini et l'agent ne possède rien. Mais comme l'agent possède des concepts, alors certains concepts au moins ne sont pas appris. Sous cette forme, l'innéisme des concepts implique de nativisme [Fodor, 1998, McDowell, 2002]. L'apprentissage de nouveaux concepts – c'est-à-dire de concepts non innés – se réduit alors à la compositionnalité du concept. Ainsi un agent peut apprendre le concept BOUTON DE MANCHETTE à partir des concepts BOUTON et MANCHETTE suivant la fonction de composition (2.10) énoncée précédemment. Suivant cette conception, le concept appris peut se réduire aux concepts primitifs qui le composent, et si ces concepts sont innés, c'est-à-dire déjà possédés par l'agent, alors l'agent n'enrichit pas son vocabulaire conceptuel. Cependant, et contrairement à Fodor, si la composition ne concerne pas le contenu mais la désignation de ce contenu, c'est-à-dire que la composition ne s'opère qu'au niveau de la fonction caractéristique, alors le nouveau concept a une application et un usage propre. Applications et usages qui ne dépendent pas des composants, mais de l'ensemble des concepts possédés par l'agent. Fodor s'opposerait évidemment à cette forme de holisme.

Cependant si le concept appris à partir d'une composition de concepts est un concept, il possède alors ses propres conditions d'application et d'usage ou bien alors est redondant avec un concept déjà possédé et alors ne sera pas utilisé et, par économie, sera ôté du vocabulaire conceptuel de l'agent (cf. chapitre 6). L'application et l'usage sont alors appris par l'agent. Un agent peut parfaitement former le concept BOUTON DE MANCHETTE et ne pas savoir à quoi ce concept réfère et comment il s'emploie, l'agent, alors, ne sera pas dit maîtriser le concept. L'application fait appel à l'extension du concept, c'est-à-dire à la référence cognitive ou canonique, postulée à partir de la fonction caractéristique. La composition de concept n'est donc pas suffisante pour rendre compte de l'apprentissage du concept même si elle peut y participer.

Une autre manière d'expliquer l'acquisition du concept est l'apprentissage par expérience. Il ne s'agit plus alors de soutenir que certains concepts sont innés mais qu'ils sont *primitifs*, c'est-à-dire qu'ils peuvent être possédés directement, sans médiation d'un autre concept. [Locke](#) et les empiristes contemporains comme [Prinz](#) soutiennent que l'expérience seule permet d'acquérir des concepts. Les données de la perception sont agrégées en ensemble et à partir d'une certaine masse critique, un label est associé à l'ensemble qui devient alors un concept. Cette méthode d'acquisition de label d'ensemble est celle utilisée en intelligence artificielle pour l'apprentissage automatique [[Cornuéjols and Miclet, 2003](#)]. Une machine équipée de capteurs, d'un système de traitement du signal et d'une mémoire, est capable de constituer des classes de stimuli. Cependant, même si l'expérience joue ou peut jouer un rôle fondamental dans la constitution du concept, elle n'est pas suffisante non plus. Imaginons une machine capable de classer des échantillons sonores sur la simple base de la fréquence du signal perçu. Une telle machine peut, après un nombre conséquent de stimulation, établir une carte d'échantillons sonores, très fine [[Sylvand, 2002](#)] mais s'avère incapable de faire quoi que ce soit de cette indexation sans indication extérieure. Par exemple, si la machine a regroupé ensemble tous les sons produits par une trompette, elle est incapable de dire qu'il y a un point morphologique commun entre ces sons et ceux produits par un trombone, combien même par ailleurs la machine aurait créé trois classes, une regroupant les sons de trompette, la seconde, les sons de trombone et la troisième les sons de trompette et de trombone. Le problème est que toutes les classes sont au même niveau, et donc qu'aucune inférence de classes en classes n'est possible. Or ces inférences sont nécessaires pour déterminer l'usage du concept. Ce problème est celui que rencontre la théorie empiriste de [Prinz \[2002\]](#). Le proxytype correspond tout au plus à un percept, c'est-à-dire à une propriété saillante de la perception, mais pas à un concept.

Que l'expérience joue un rôle crucial dans l'apprentissage du concept, ou du moins de certains types de concepts, cela semble évident. Il est difficile d'imaginer que les concepts perceptuels soient complètement séparés de toute expérience. Cependant, si le contenu ou l'application peut passer par l'expérience, l'usage du concept dépend de la relation que ce concept entretient avec les autres concepts et états mentaux dont l'agent dispose.

Il est possible d'imaginer un scénario dans lequel un agent apprendrait à discriminer et à reconnaître la couleur rouge par exemple, ce qui signifierait au moins qu'il possède le percept 'rouge', voire le concept s'il est en mesure d'identifier les instances de cette couleur. L'*identification*, contrairement à la *discrimination*, requiert en effet le concept, dans la mesure où elle est une représentation satisfaisant le principe de correction. Alors que la discrimination implique de reconnaître que deux propriétés dans un mode de présentation sont différentes, suivant un *principe contrastif*, l'identité suppose s'assigner une propriété à un objet alors qu'il apparaît sans contraste avec aucun autre dans un mode de présentation.

Unless a creature has the resources to make contrastive identification and/or comparisons among different instantiations of one and the same visual attribute or property, she will not be able to recognize or re-identify the property or attribute in question ¹⁰.
[Dokic, 2002]

$$(\text{Principe d'identité}) \quad \forall z(x = z):(\lambda zx = \lambda zy) \quad (2.11)$$

Si le contexte de l'expérience ne requiert pas nécessairement de concept, du fait que le traitement de l'information des données sensorielles s'effectue au niveau non-conceptuel, à travers la sélection de propriétés saillantes et de proto-jugements, l'expérience ne peut fournir que des percepts ou ce que Peacocke appelle des proto-concepts ou ce que Prinz nomme des proxytypes. Comme dans le cas de la composition des concepts, ces percepts doivent être individués, au sens fort de l'identification, à partir de leurs traits fondamentaux, c'est-à-dire de leur fonction caractéristique. La profondeur de la détermination de la fonction caractéristique, ce que l'on appelle également le « grain » de la description, dépendant du nombre de fonctions caractéristiques considérées. L'identification des percepts se fait à l'aide d'un abstrait correspondant aux propriétés saillantes déterminant les percepts. Le passage à l'abstrait [Church, 1956] conduit à la représentation et à l'épistémicité tels que définis précédemment, c'est-à-dire au niveau concep-

¹⁰ «À moins qu'une créature ait les ressources pour faire une identification contrastive et/ou des comparaisons parmi différentes instantiations d'un et du même attribut ou propriété visuelle, elle ne sera pas capable de reconnaître ou de ré-identifier la propriété ou l'attribut en question.»

tuel [Lewis, 1970a, McDowell, 2002]. L'abstrait est requis pour permettre la catégorisation et la combinaison. Le concept doit permettre de subsumer n'importe quel objet possédant les propriétés fondamentales qui apparaissent dans la fonction caractéristique du concept, et de ce fait le concept ne doit pas dépendre, pour son application ou son usage, d'aucune de ses instances. Des auteurs comme McDowell ou Noë affirment que la conceptualisation implique que le contenu de l'expérience contient déjà les concepts qui permettent de la décrire, sans cela aucune connaissance n'est possible [Lewis, 1970a]. En somme ces auteurs refusent toute notion de contenu non-conceptuel de l'expérience. Or il semble qu'il y ait bien quelque chose comme un tel contenu (cf. section 2.1). Cependant le concept pourrait être généré à partir d'un contenu de perception, contenu non-conceptuel mais structuré en percepts c'est-à-dire en propriétés saillantes auxquelles l'agent peut porter attention sans en avoir de connaissance, et d'un traitement de cette information à partir d'une capacité à faire des inférences, c'est-à-dire une sorte de « module de la logique » – logique cognitive et non pas mathématique évidemment.

L'apprentissage artificiel, en intelligence artificielle, ne donne des résultats probants que s'il est supervisé. Un apprentissage est « supervisé » lorsqu'à un robot doté de capteur sensoriel et d'un système de traitement du signal tel qu'il est capable de former de lui-même des ensembles d'échantillon similaire, est donnée une liste de descripteurs avec des caractéristiques associées. Pour reprendre l'exemple des instruments de musique, on peut donner un descripteur « cuivre » par exemple, et indiquer que les échantillons avec telle morphologie (avec un certain timbre et une certaine fréquence par exemple) sont produits par les cuivres. Le robot confronte alors ses propres classes avec les descripteurs donnés et revisite celle-ci à partir des contraintes de ceux-là. Il en résulte une classification automatique très performante, et ce principe n'est autre que celui utilisé dans les bases de données dynamiques. Cependant, il faut donner des descripteurs à la machine, et ces descripteurs font office de concept, et c'est en ce sens que les opposants au contenu non-conceptuel soutiennent qu'un apprentissage non supervisé, c'est-à-dire sans concept, n'est pas possible.

Une autre manière de superviser l'apprentissage ne serait pas de donner des descripteurs mais des relations entre ensembles. Ces *connecteurs* n'auraient aucun contenu, mais permettraient de relier des ensembles entre eux

de manière à pouvoir former des inférences. Des inférences apparaissent au niveau non-conceptuel sous forme de réflexes et de routines, comme l'illusion de persistance de taille le montre (cf. section 2.1). Ces connecteurs ne seraient donc pas conceptuels, mais utilisés par le niveau conceptuel.

2.3.5 Publicité du concept

L'apprentissage du concept tel qu'il est exposé dans la section précédente, correspond en fait à l'acquisition de concept. Une autre manière d'apprendre un concept passe par la communication. Si Paul pense que les baleines sont des poissons mais que Pascal lui dit que ce sont des mammifères, Paul peut corriger son concept, ou bien si Paul demande ce que sont des animaux qui nagent là-bas, et que Pascal lui dit que ce sont des baleines, Paul peut apprendre ce concept. Mais pour que cela soit possible, il faut qu'une communication soit possible entre agents cognitifs. La communication passe pas un code de signes qui véhiculent un sens. Le terme lexical 'baleine' réfère à ce qui désigne les baleines dans le monde, c'est-à-dire le concept BALEINE. Le concept est différent du terme lexical, puisque deux termes lexicaux différents peuvent désigner un même concept, il suffit de penser à la traduction de langue à langue par exemple, le terme 'baleine' en français et le terme 'whale' en anglais désignent tous deux le concept BALEINE.

La communication suppose le partage de concepts entre agents cognitifs : le concept doit donc être public [Fodor, 1998, p. 28].

Un concept est public s'il peut être possédé par deux agents cognitifs distincts. Paul et Pascal partagent tous deux le concept BALEINE si dans chacun des espaces cognitifs réciproques de Paul et de Pascal apparaît ce concept.

Le problème est que l'application et l'usage du concept dépendent des autres concepts possédés par l'agent et de l'expérience que celui-ci a de l'environnement. Un agent qui aura déjà vu des baleines et qui est spécialiste de faune marine n'aura probablement pas les mêmes pensées, et donc le même concept, qu'un agent qui n'en aura jamais vu. L'exemple de Paul et de Pascal, où Paul pense que les baleines sont des poissons et Pascal pense que ce sont des mammifères révèle que le concept BALEINE est différent dans l'espace cognitif de l'un et de l'autre. Comment comparer ces concepts ?

L'énoncé :

Les baleines sont des poissons (2.12)

sera tenu pour vrai par Paul et faux par Pascal. L'application du concept pour l'un et pour l'autre pourrait être équivalente – Paul et Pascal pourraient désigner les mêmes objets de l'environnement en pensant que ce sont des baleines – mais les inférences qu'ils tirent de ces concepts sont différentes, donc l'usage du concept diffère d'un espace à l'autre. Intensionnellement ces deux concepts seront donc distincts. Cependant Pascal peut corriger le concept de Paul, en lui apprenant que les baleines sont des mammifères et non pas des poissons, mais pour cela il faut que Pascal comprenne que Paul souhaite parler des baleines et non pas d'autre chose. Imaginons que Paul pense que le terme lexical 'baleine' s'applique aux daurades, alors l'énoncé ci-dessus est vrai, mais le terme 'baleine' n'a pas sa signification normale.

La difficulté de la publicité du concept est qu'elle passe par la communication, et que la communication est régie par des normes et des règles. Cette communication n'est pas nécessairement verbale une action ou une gestuelle peut servir à communiquer une pensée et donc véhiculer des concepts.

Il faut distinguer le terme lexical du concept, tout comme l'application et l'utilisation du terme lexical de celles du concept [Sperber and Wilson, 1986].

La question de la publicité du concept rejoint la question de la « division du travail linguistique » soulevée par Putnam [1975], et la manière dont deux agents cognitifs d'une même communication linguistique partagent l'usage et la référence d'un terme lexical. Lorsqu'une dame se plaint de son arthrite en se frottant la cuisse, ce qu'elle veut exprimer, c'est-à-dire ce qu'elle pense, est qu'elle a mal à la cuisse, et que sa maladie a un nom et qu'elle pense que le terme 'arthrite' s'applique à la maladie qui lui cause cette douleur. Or cette dame ne peut souffrir d'arthrite dans la cuisse pour la simple raison que l'arthrite est une maladie articulaire et qu'il n'y a pas d'articulation dans la cuisse. Un médecin peut comprendre et *postuler* ce que la dame essaie d'exprimer, à savoir une douleur musculaire, et constater le mésusage du terme lexical pour exprimer cette pensée. S'il s'agit effectivement d'une pensée que la dame essaie de communiquer, cela suppose qu'elle sait ce qu'elle veut exprimer – une douleur à la cuisse – et cette pensée est composée

de concepts. Dire que la dame sait ce qu'elle souhaite exprimer ne signifie pas qu'elle sache qu'elle est atteinte de myosis ou même qu'elle sait que la partie anatomique qu'elle désigne s'appelle la cuisse, mais elle a un accès à son contenu mental au sens donné au épistémique donné plus avant (cf. section 2.2.2).

Le terme lexical 'arthrite' a un usage et une application définie dans la communauté linguistique dans laquelle ce terme apparaît. Cet usage et cette application sont régis par une norme qui peut être explicitée au besoin par un groupe d'experts reconnus comme tels au sein de cette communauté. Un médecin peut expliquer ce qu'est l'arthrite à sa patiente, et au besoin lui montrer la définition de ce terme dans le Vidal, livre de référence. Mais si une convention peut se comprendre au niveau du terme lexical, qu'en est-il au niveau de la pensée. Y a-t-il une convention, une norme ou une « division du travail conceptuel » au niveau de la pensée, c'est-à-dire au niveau du concept ?

L'innéisme a l'avantage de supposer que les concepts – certains au moins – font partie de l'appareillage cognitif, de sorte que pour être cognitif, un agent doit déjà posséder ces concepts. L'innéisme va de pair avec une sémantique causale de la référence du concept. Comme la possession d'un concept ne suppose pas la possession d'autres concepts par l'agent, les relations entre les concepts de l'agent et leur application n'interfère pas avec l'application et l'usage d'un concept particulier. De fait, si un agent possède le concept BALEINE il possède également de facto ses conditions d'usage et d'application. Suivant cette conception, défendue par l'atomisme conceptuel de Fodor, Paul ne possède pas le concept de baleine s'il pense que ce sont des poissons, mais le concept PALEINE, et une fois qu'il aura accepté la remarque de Pascal comme quoi les baleines sont des mammifères, alors son concept PALEINE sera remplacé par le concept BALEINE. Le seul inconvénient est que l'on ne comprend pas vraiment comment Pascal et Paul peuvent savoir qu'ils essayent de parler de la même chose, à savoir de baleines.

Deux agents comprendront que la communication entre eux échoue lorsque l'un admet quelque chose de vrai que l'autre tient pour faux. C'est-à-dire à partir du moment où une inférence à partir d'un concept est permise dans la conception de l'un des agents mais pas dans celle de l'autre, ou bien quand les conditions d'applications du concept sont différentes.

Une manière de vérifier que deux agents pensent à la même chose dans leur communication serait d'avoir recours à une norme, à un dictionnaire de concepts comme dans le cas de la communication lexicale. Le problème est qu'il n'existe rien de tel. Il pourrait exister une sorte de division du travail conceptuel, à l'image de la division du travail linguistique. L'application et l'usage de référence du concept ARTHRITE serait ceux possédés par des experts. Le problème est que dans le cadre de la division du travail linguistique, la norme se fait à partir de la communication entre les experts et l'échange de preuves et de raisons. Mais c'est justement cela que nous cherchons à examiner pour le concept. D'autre part quel statut aurait ce concept commun ? Serait-il intersubjectif ? objectif ? un agent posséderait-il son propre concept plus ce concept commun, ou bien seulement le concept commun ? S'il possède les deux, lequel doit-il utiliser ? S'il ne possède que le concept commun comment peut-il se tromper ?

Lorsque médecin explique à la dame qu'elle ne peut pas être atteinte d'arthrite, puisqu'elle se plaint d'un mal à la cuisse, il imagine que ce que pense la dame est quelque chose comme « j'ai mal à la cuisse et je crois que c'est à cause de l'arthrite », et ce qu'essaie de dire le médecin est quelque chose comme « oui vous avez mal à la cuisse, mais cette maladie ne s'appelle par 'arthrite' mais 'myosis' ». En somme il postule la pensée de la dame, et essayant de postuler le sens que la dame associe au terme 'arthrite', conserve ce sens et l'applique à un autre terme. Il ne va de même pour Paul et Pascal. Pascal essaye de penser ce que Paul pense par le terme 'baleine' et essaie de l'adapter au sens de son propre terme. Le médecin possède donc son propre concept, ainsi que le concept qu'il imagine être celui de la dame, confronte les deux et les restitue. Il y a donc un concept « privé », *cognitif* et un concept « postulé », *canonique* (cf. chapitre 4). L'agent possède son propre concept mais pour ce concept en plus de l'application et de l'usage qu'il en fait – applications et usages qui correspondent à la référence et à la dérivation cognitive, c'est-à-dire propre à l'agent suivant son expérience et les concepts qu'il possède – l'agent *postule* une référence et une dérivation qu'aurait, selon lui, un autre agent cognitif pour ce même concept. Cette référence et la dérivation postulée est généralement identique à la référence et à la dérivation cognitive, sauf dans les cas de divergence et d'échec de la communication.

La référence et l'extension postulée par l'agent ont deux niveaux : l'un est objectif et correspond au concept d'un *expert*, l'autre est *intersubjectif* et correspond au *sens commun*. Le sens commun se traduit, au niveau d'une communauté à travers les théories naïves que celle-ci admet. Le sens commun forme la conception admise de manière canonique mais non nécessairement scientifique, par une communauté. À travers ces conceptions, il est possible de comprendre les concepts dans leur acception intersubjective.

Chapitre 3

Conceptions du concept

3.1 Théorie classique

Théorie classique : *les concepts sont des définitions de traits caractéristiques nécessaires et suffisants pour catégoriser les objets.*

La « théorie classique » [Smith and Medin, 1981] soutient que le concept est une *définition*, c'est-à-dire une liste de caractères qu'un objet doit nécessairement posséder pour être subsumé par un concept et que ces propriétés sont suffisantes pour catégoriser un objet. La fonction principale du concept est alors la classification, et la structure du concept respecte le principe du tiers exclu de la logique classique, c'est-à-dire qu'un objet tombe sous une et une seule catégorie de même niveau, et que tout objet appartient ou n'appartient pas à une catégorie donnée. De fait tout objet appartient nécessairement à une catégorie, sinon il n'apparaît pas comme objet et alors le problème ne se pose pas. L'engagement épistémique demandé par cette conception à l'agent cognitif est fort, en effet l'agent qui possède le concept doit être en mesure de dire de quelle catégorie relève un objet.

La théorie classique est appelée ainsi puisqu'elle est censée être celle défendue par Platon, Locke ou encore Frege [Smith and Medin, 1981].

Thus, the idea of the sun, what is it but an aggregate of those several simple ideas, bright, hot, roundish, having a constant regular motion, at a certain distance from us, and perhaps some other : as he who thinks and discourses of the sun has been

more or less accurate in observing those sensible qualities, ideas, or properties, which are in that thing which he calls the sun ¹.

[Locke, 1997, II, 23, §6]

En fait l'εἶδος de Platon ne peut pas correspondre à proprement parler au concept dans son acception actuelle, puis que l'Idée est indépendante de l'agent et n'est pas cognitive pour Platon. L'agent certes saisi l'Idée en percevant la forme dans la matière, mais ce qui se passe dans l'espace cognitif de l'agent est distinct de l'Idée. De fait les caractéristiques de la forme sont certes nécessaires et suffisantes mais son rôle n'est pas de constituer la saisie et la représentation de la chose, mais la chose elle-même, ce qui n'est pas tout à fait la même chose. Quand à l'*Idée* au sens moderne donné à ce terme par Locke elle n'est pas exactement le concept non plus bien qu'elle s'en rapproche. À la différence de Descartes et de Platon, Locke affirme que la relation entre l'agent et le monde modifie la représentation que le premier à du second. C'est à travers l'expérience sensible que l'agent va se forger une représentation de l'environnement, en sélectionnant des propriétés perçues et en les agrégeant pour en former des unités de connaissances : les Idées. Les caractères nécessaires et suffisants de l'idée dépendent donc de l'agent, de l'expérience qu'il a du monde et des idées qu'il possède déjà. En ce sens, la théorie des Idées de Locke se rapproche plus de la théorie empirique du concept ou de la sémantique des rôles conceptuels que de la théorie classique.

Si les caractères du concept sont censés être nécessaires et suffisants pour déterminer l'appartenance ou non d'un élément à une catégorie, la théorie classique ne dit pas sur quelle base ces caractères sont déterminés. Pour Platon les caractères sont ceux qui font que l'élément est ce qu'il est, le critère est ontologique. Pour Locke il s'agit des qualités perçues. Mais nous venons de dire que ni la conception platonicienne ni la conception lockienne ne correspondent exactement au concept. Quant à Frege sa notion du concept est complexe à saisir. En effet le concept est un abstrait indépendant de l'espace cognitif de l'agent – c'est son anti-psychologisme – qui est pourtant saisi par l'agent lorsqu'il exprime une pensée. Le concept est le *sens* de la

¹“Ainsi, l'idée du soleil, n'est qu'un agrégat de ces quelques idées simples, lumineux, chaud, rondeur, avoir un mouvement constant et régulier, être à une certaine distance de nous, et peut-être d'autres : comme qui pense et discute du soleil a été plus ou moins précis dans l'observation de ces qualités sensible, idées ou propriétés qui sont dans cette chose qu'il appelle le soleil.”

pensée, et les caractères nécessaires et suffisants n'y entrent que de manière indirecte en dépendant de la connaissance que l'agent a de l'objet de sa pensée. Ainsi il peut considérer que le vainqueur de Iéna et le vaincu de Waterloo sont deux personnes différentes, et donc penser que ce sont des caractères nécessaires et suffisants, parce qu'il ne sait pas qu'il s'agit en fait de Napoléon dans les deux cas.

Si la détermination des caractères se fait sur une base ontologique, alors le concept n'est plus une entité mentale. Si elle se base sur la compréhension du contenu de l'expérience alors elle n'est pas suffisante pour déterminer l'appartenance d'un objet à une catégorie. Si elle dépend de la connaissance de l'agent alors elle n'est pas nécessaire. La pertinence du critère dépend des concepts possédés par un agent. Si l'agent possède simplement deux concepts ANIMÉ et INANIMÉ, alors le concept ANIMÉ est nécessaire et suffisant pour caractériser un chien par exemple. Mais si l'agent possède plusieurs concepts, comme ceux de CHAT, COLORÉ, POSSÉDÉ PAR PAUL, alors le simple fait de dire qu'un chien est animé n'est pas suffisant pour caractériser un objet du monde, puisque les chats sont animés sans être des chiens.

Le concept, selon la théorie classique est une définition de la forme suivante :

x relève du concept α si et seulement s'il possède les caractères $\Phi\{c_1, \dots, c_n\}$.

de sorte que le concept α est une fonction d'assignation définie sur deux valeurs selon que l'objet x possède ou ne possède pas les caractères en question.

$$\alpha \mapsto 0, 1 \tag{3.1}$$

tel que

$$f : x \mapsto \alpha \begin{cases} 1 & \text{si et seulement si } x \text{ possède } \Phi \\ 0 & \text{sinon} \end{cases} \tag{3.2}$$

Cette fonction est la fonction de [catégorisation](#).

Les caractères Φ devant être nécessaires et suffisants, tout objet doit pouvoir être catégorisé dans une classe et une seule de même niveau :

$$\Box \forall \alpha \forall x (x \in \alpha \vee x \notin \alpha) \quad (3.3)$$

À quel niveau cette assignation doit-elle être comprise ? Est-ce que cela signifie que tout élément est déterminé de manière absolue – au sens ontologique platonicien du terme – ou bien que l’agent cognitif doit nécessairement être en mesure de dire de quelle catégorie un élément donné relève ? S’il s’agit du point de vue ontologique alors ce n’est pas de concept dont il est question mais de *propriétés*. S’il est évident que tout élément est ce qu’il est du fait de ce qu’il est, cela ne signifie pas pour autant qu’un agent a besoin de le savoir. Si au contraire cette assignation est épistémiquement déterminée, c’est-à-dire qu’un agent doit être en mesure de dire pour tout élément quelle est sa nature, alors cela implique que le sujet soit omniscient et qu’aucun de ses états mentaux n’interfère dans l’évaluation des caractéristiques de l’élément. Un conceptualisme tel que celui défendu par Lewis [1970a] , McDowell ou Noë laisse supposer que l’engagement épistémique fort est requis pour toute interaction avec le monde. Or l’analyse du problème de la persistance de l’illusion montre que ce n’est pas le cas. L’agent peut inter-agir avec l’environnement en utilisant des caractéristiques de l’élément – ses propriétés saillantes – sans nécessairement utiliser de concept.

L’avantage de l’approche du concept comme définition est que cette théorie des concepts explique très bien l’aspect catégoriel du concept. Un objet tombe sous un concept et un seul de même niveau du fait des propriétés qu’il possède. La structure de la classification est arborescente – taxinomique – du fait du strict respect du tiers exclu. La théorie classique rend alors compte de la hiérarchie des concepts, c’est-à-dire du fait que le concept BALEINE est inclu dans le concept MAMMIFÈRE, et donc qu’un élément qui est une baleine est nécessairement un mammifère, mais la réciproque n’est pas nécessairement vraie. Si Φ est Ψ alors Ψ apparaît dans la liste des propriétés de Φ . De cette hiérarchie découle la transitivité des propriétés. Si tous les Φ sont des Ψ et que tous les Ψ sont des Θ alors tous les Φ sont des Θ . Puisque par définition, Θ doit être inclus dans Ψ qui doit être inclus dans

Φ , ce qui implique que Θ doit nécessairement apparaître dans Φ en tant que propriété. En expliquant ainsi la hiérarchie des concepts, la théorie classique explique également comment il est possible d'apprendre un concept à partir d'un autre, comment il est possible d'apprendre Ψ à partir de Φ ou de Θ .

Cette structure ne reposant que sur les propriétés des concepts, c'est-à-dire leur fonction caractéristique il est possible de définir n'importe quel concept, puisque les abstraits peuvent se définir [Hilbert and Bernays, 2001a].

D'autre part, l'appartenance de l'élément au concept étant déterminé par ses caractères propres, l'appartenance n'est pas définie en terme de ressemblance. La table sur laquelle j'écris n'est pas une table parce qu'elle ressemble à celle qui est dans la cuisine, mais elle est une table parce qu'elle a les propriétés qui font d'elle une table, et si elle ressemble à celle qui est dans la cuisine c'est que celle-là possède ces caractères aussi. L'appartenance à un concept est donc un problème de relation entre l'élément et le concept, et non pas entre les éléments. De sorte qu'un xoloitzcuintle sera tout aussi un chien qu'un caniche, et un loup ne sera jamais un chien. La théorie classique, en se conformant au tiers exclu est très à l'aise avec la logique classique et les concepts peuvent ainsi être considérés comme des ensembles ou des prédicats.

La théorie classique explique la distinction de concepts co-extensionnels à l'aide des différences qui apparaissent dans les fonctions caractéristiques. Les concepts VAINQUEUR DE IÉNA et VAINCU DE WATERLOO désignent tous les deux Napoléon, mais décrivent l'élément Napoléon de manière différente, si bien qu'il n'est pas impossible d'imaginer un environnement, un monde possible dans lequel ces deux concepts référerait à des objets distincts. Ces deux concepts sont donc différents.

Cependant, ce qui apparaît comme des avantages de la théorie classique se retrouve également être des faiblesses. Bien sûr qu'un xoloitzcuintle est un chien tout comme un caniche, mais il n'en reste pas moins vrai que généralement, lorsque l'on présente à un agent cognitif trois photographies, l'une d'un xoloitzcuintle, une autre d'un caniche et la troisième d'un loup, et qu'on lui demande quels sont ceux qui sont des chiens, il n'hésite pas pour le caniche, peu pour le loup et ne sait pas pour le xoloitzcuintle. Cette hésitation provient de l'*effet de typicité* [Rosch, 1975] : certains membres

d'un concept apparaissent comme étant de meilleurs représentants de ce concept que d'autres. L'effet de typicité indique qu'un agent cognitif n'accorde pas le même poids aux différentes instances d'un même concept. Certains éléments de l'extension d'un concept semblent représenter plus le concept que d'autres. Par exemple, lorsque dans un roman il est dit quelque chose comme « le chien se mit à aboyer puis sauta sur le sofa », sans plus d'indications sur le pedigree de l'animal et que l'on demande à un agent cognitif à quoi, selon lui, ressemble ce chien, la proportion de réponses affirmant que c'est un caniche est largement supérieure à celles disant que c'est un xoloitzcuintle, tout comme un moineau est jugé être plus un oiseau qu'un kiwi, ou une pomme plus un fruit qu'une tomate. Cela signifie que l'agent sélectionne certaines propriétés nécessaires et suffisantes déterminant l'appartenance d'un élément à un concept, comme étant meilleures ou plus importantes que d'autres. Or cela la théorie classique ne le dit pas. Elle prédit même le contraire. Une tomate est un fruit parce que les tomates ont tout pour faire d'elles des fruits. Et les tomates n'ont pas moins de ces propriétés essentielles que n'en ont les pommes. S'il est vrai que les éléments doivent posséder des propriétés essentielles, nécessaires et suffisantes pour déterminer de quel concept ils relèvent, cela semble nécessaire pour qualifier ontologiquement l'élément, mais pas pour en rendre compte cognitivement, c'est-à-dire de la manière dont un agent va le représenter.

Mais le problème n'est pas dans le fait que l'agent cognitif ne semble pas utiliser uniquement – si jamais – la définition du concept pour l'appliquer, le problème est également qu'il est difficile de trouver une définition, c'est-à-dire une liste de caractères nécessaires et suffisants, pour déterminer le concept. Quelles sont les propriétés nécessaires et suffisantes qu'un élément doit avoir pour être un oiseau, c'est-à-dire pour être subsumé par le concept OISEAU ? Avoir des ailes ? mais qu'en est-il du Kiwi ? Avoir des plumes ? Mais qu'en est-il du manchot ? Pour chaque caractère listé quasiment, une exception peut-être trouvée. À moins que l'on utilise les critères taxinomiques. Mais il est fort peu probable que tous les agents cognitifs possédant le concept OISEAU connaissent et utilisent ces critères-là. D'autant plus que ces critères dépendent en fait d'une théorie scientifique, et que les critères de la cladistique ne sont pas nécessairement les mêmes que ceux de la systématique, si

bien que selon la théorie adoptée, un oiseau n'aura pas les mêmes caractères essentiels.

3.2 Théorie-théorie

Théorie Théorie : *les concepts sont structurés comme des théories, avec des axiomes, des théorèmes et des règles de déduction.*

La théorie classique des concepts soutient que les concepts sont des définitions incluant la liste des caractères nécessaires et suffisants pour déterminer l'appartenance d'un élément à un concept, définition que l'agent cognitif doit connaître et appliquer pour posséder et utiliser un concept. Mais outre le fait qu'il soit difficile – pour ne pas dire impossible de dresser la liste de ces caractères nécessaires et suffisants, les expériences de psychologie montrent que l'agent n'utilise pas ces définitions – ou tout du moins pas uniquement les définitions – pour appliquer le concept.

Afin d'expliquer le caractère « relatif » des définitions, la *théorie-théorie* des concepts (*Theory Theory of concepts*), décrit le concept comme étant une théorie similaire à une théorie scientifique.

La théorie des concepts comme théorie se base sur une interprétation forte du caractère épistémique du concept.

Concepts represent our knowledge of the kinds of things in the world ². [Murphy, 2002, p. 392]

Le concept est donc une connaissance. L'agent qui le possède sait non seulement qu'il le possède mais connaît également les conditions d'usage et d'application du concept. Cette forme de concept comme connaissance est sous-tendue parce que l'on pourrait appeler le « conceptualisme », une théorie selon laquelle l'agent ne pourrait pas interagir avec l'environnement s'il n'avait pas connaissance ni de celui-ci ni de son interaction avec. Le conceptualisme sous cette forme est défendu par des auteurs tels que Brandom, McDowell ou encore Noë. De fait, tout contenu non-conceptuel est nié.

L'argument général se retrouve dans une critique du nominalisme par Lewis [1970a] :

²“Les concepts représentent notre connaissance des types de choses dans le monde.”

1. That which is or can be the same in two instances is abstract.
2. Nothing which is not the same in two instances can be recognized.
3. That which is not recognizable is not identifiable.
4. Hence if any non-abstract entities are real they cannot be identified.
5. What cannot be identified cannot be known.
6. According to nominalism, only non-abstract entities are real.
7. Hence, admitting nominalism, if anything is real it cannot be known.
8. Hence the only consistente nominalist is one who is also a skeptic.

Skeptics should not make statements about the character and limits of the real, since they admit that all such statements are unprovable. Perhaps, thesis (4) is a critical member of the set. If so, the point would not be that non-nominalists admit (4) but that they have a way of obviating it; and the question may be whether any consistent theory of knowledge, obviating (4), is open to the nominalist ³.

³ «

1. Ce qui est ou peut être le même dans deux instances est abstrait.
2. Rien qui n'est le même dans deux instances ne peut être reconnu.
3. Ce qui ne peut être reconnu ne peut être identifié.
4. Par conséquent si des entités non abstraites sont réelles, elles ne peuvent être identifiées.
5. Ce qui ne peut être identifié ne peut être connu.
6. Selon le nominalisme, seules les entités non abstraites sont réelles.
7. Par conséquent, suivant le nominalisme, si quelque chose est réel cela ne peut être connu.
8. Par conséquent, le seul nominaliste consistant est également un sceptique.

. Les sceptiques ne devraient pas faire des énoncés au sujet du caractère et des limites du réel, puisqu'ils admettent que tous les énoncés de ce type ne sont pas prouvables. Peut-être que la thèse (4) est critiquable dans cet ensemble. Si c'est le cas, le point ne serait pas que les non nominalistes admettent (4) mais qu'ils ont un moyen de l'éviter; et la question est peut-être celle de savoir si toute théorie consistante de la connaissance, évitant (4), est accessible au nominaliste.”

Posséder un concept c'est être capable de l'appliquer, c'est être donc capable de dire que deux éléments sont des instances d'un même concept. Par conséquent le concept est commun à deux particuliers, ce qui fait de lui un *terme général*, c'est-à-dire un *universel* ou un *abstrait*. Seul ce qui est commun à plusieurs particuliers peut être reconnu, c'est-à-dire que seuls les caractères sont utilisés par la reconnaissance, pour la simple raison que seul ces caractères peuvent être identifiés, suivant le principe d'identité (2.11). De fait, seuls les termes abstraits peuvent être identifiés. Comme la reconnaissance ne porte que sur les abstraits, seuls les abstraits peuvent être connus.

Cette critique, dirigée contre le nominalisme, implique que le concept est ce qui est commun à plusieurs particuliers – ses instances – et que l'identification des éléments passe nécessairement par la saisie du concept qui, en tant qu'il permet la reconnaissance, permet la connaissance. Cet argument se retrouve chez McDowell [2002] ou Noë [2004], il consiste à dire que pour pouvoir être utilisée, l'expérience sensible doit être comprise par l'agent. Or pour pouvoir être comprise, l'expérience doit nécessairement déjà être conceptuelle.

Mais comme la définition du concept en terme de définition ne marche pas, l'idée est de dire que les concepts sont comme des théories scientifiques : qu'ils sont composés d'axiomes et de théorèmes, d'hypothèses et de protocoles, mais au lieu d'être scientifique, ce sont des *théories naïves* (*folk theory* – *naive theory*). Une théorie naïve est une théorie naturellement possédée par un agent cognitif du fait de son interaction avec le monde. Il y a une physique naïve qui dit que lorsque j'allume le gaz sous la bouilloire remplie d'eau, l'eau au bout d'un moment devient assez chaude pour pouvoir faire du thé. L'explication de l'échauffement de l'eau, est la flamme du gaz. Mais c'est là une explication naïve – c'est-à-dire une explication qui convainc l'agent cognitif, mais qui ne repose pas sur des bases scientifiques. Et dans certains cas, j'allume le gaz sous la bouilloire pleine d'eau et l'eau, au bout d'un moment est toujours aussi froide et je ne comprends pas pourquoi. L'explication scientifique dit alors que c'est dû au manque d'oxygène ou au fait qu'il fait -40 °C dehors. Tout comme il y a une physique naïve, il y a une biologie naïve, une psychologie naïve et toute autre explication naïve de ce que l'agent a besoin de comprendre [Atran, 1990, Povinelli et al., 2003, Bloom, 2004]. Ces théories naïves sont très robustes et constituent une base

de préjugés difficile à contrecarrer. Par exemple, imaginons que je vous dise que je viens vous dire que j'ai payé 11 € une raquette de tennis et une balle. La raquette m'a coûté 10 € de plus que la balle. Selon l'économie naïve, je suis à peu près certain que si je demande combien m'a coûté la balle, la prédiction est que la plupart des agents cognitifs répondront que la balle a coûté 1 €, alors même qu'elle n'a coûté que 30 centimes. Les théories naïves sont incomplètes – l'astronomie naïve ne dit pas grand-chose des quasars – mais elles fournissent des explications satisfaisantes pour la curiosité de l'agent et sont suffisantes pour lui permettre d'interagir dans la plupart des cas, avec son environnement.

Si les concepts sont structurés en théories naïves cela signifie qu'ils sont reliés les uns les autres. D'ailleurs il n'est pas clair de savoir à quel niveau le concept est une théorie : est-il une théorie en lui-même ou bien fait-il parti d'une théorie. Quoi qu'il en soit, si le concept est ou fait parti d'une théorie, même naïve, cela suppose qu'il est dépendant d'autres concepts possédés par l'agent. Par exemple le concept OURS POLAIRE est dépendant des concepts DOUX et MÉCHANT. Les théories naïves dessinent ainsi des domaines d'application et d'usage spécialisés, ces domaines étant des trames de relations spécifiques entre des concepts. En d'autres termes, la théorie des concepts comme théories, tout comme la théorie classique des concepts, suppose une forme de holisme, au moins partiel [Quine, 1953]. Ce holisme permet à la théorie classique d'expliquer un concept à l'aide d'un autre ou d'autres concepts, par exemple qu'un célibataire est une personne non mariée, et à la théorie des concepts, et d'expliquer les domaines d'application des concepts, c'est-à-dire la physique naïve, la biologie naïve, et les autres théories naïves utilisées par l'agent cognitif. La relation entre concepts est inférentielle, elle détermine le passage d'un concept à un autre [Margolis and Laurence, 1999, p. 45].

Les partisans de la théorie des concepts, par exemple Gopnik and Meltzoff [1996] comme théorie considèrent que le concept est une théorie au même sens qu'une théorie est scientifique, c'est-à-dire comme un ensemble d'entités abstraites, de lois, d'hypothèses qui transcende l'expérience, qui est nomologique qui détermine des régularités logiques. Le concept FAIRE CHAUFFER DE L'EAU est de la forme : «si j'allume le gaz et que je mets une

bouilloire pleine d'eau dessus, alors l'eau de la bouilloire sera chaude après quelques instants».

L'avantage de cette approche est qu'elle permet d'expliquer l'apprentissage et le changement de concept, en reprenant l'apprentissage et le changement des théories scientifiques. Évidemment, si les concepts sont les théories scientifiques, il reste à expliquer l'apprentissage et le changement des théories scientifiques sans avoir recours aux concepts. Mais l'avantage est que comme le concept est l'explication pour et part l'agent de son environnement et que cette explication est théorique – même si naïve – l'agent peut faire intervenir autant de principes explicatifs qu'il juge nécessaire de le faire. Il peut avoir recours à des phénomènes non observés voire même non observables, les fameuses « propriétés cachées » (*hidden properties* [Povinelli et al., 2003]), tel que Dieu, Big Brother, la main invisible, etc. Ces entités sont fort coûteuses du point de vue ontologique, évidemment, mais si elles sont requises ou jugées telles pour l'explication, elles ne seront pas remises en cause, sauf bien sûr si l'explication n'est pas satisfaisante. D'un point de vue naïf cependant ces entités étranges sont supposées exister, du fait même qu'elles ont une incidence sur l'environnement. De sorte qu'une ontologie naïve fait partie des concepts, ontologie qui se traduit le plus souvent par une *essentialisme psychologique* [Bloom, 2004], c'est-à-dire de postuler comme réel et existant tout ce que l'on croit avoir un effet causal. Ce qui est remis en cause n'est pas que tout ce qui possède un effet causal existe, mais ce qui est reconnu comme ayant ou étant un tel effet. Ce n'est pas parce que l'on peut donner une explication de l'humeur des gens en fonction de la position de la Lune que cette explication est la bonne.

La théorie des concepts comme théories à l'avantage de coller assez bien à ce que l'agent cognitif est enclin à penser du concept, pour la simple raison que si le concept est une théorie naïve, cette théorie naïve correspond plus ou moins à l'explicitation du sens commun, l'agent se sentira réconforté si ce qu'on lui demande correspond à ce que la plupart des gens pensent. Mais dire que l'agent est d'accord avec le sens commun ne signifie pas pour autant que le sens commun est le concept. L'explication donnée plus loin dans cette étude rend compte du sens commun comme de la dérivation et de la référence canonique postulée par l'agent pour un concept qu'il possède (cf. chapitre 4). La théorie des concepts comme théories ne semble pas

distinguer le niveau de la théorie naïve du niveau du concept. Or il se trouve quelque fois que le concept de l'agent ne concorde pas nécessairement avec celui du sens commun, c'est typiquement le cas du concept de l'expert. Comment expliquer ce décalage ?

En revanche il est vrai que le concept prend place dans une théorie. Si les baleines sont des mammifères alors il y a bien une relation entre les concepts BALEINE et les MAMMIFÈRE. De fait pour apprendre un concept il suffit d'apprendre quelle est sa position dans une théorie, et cela est permis du fait de la structure holistique du concept. Les concepts sont donc individués par leurs rôles et leurs usages. Cependant cela est vrai pour les termes utilisés dans une théorie, mais pas pour la théorie elle-même, puisqu'elle dépend des termes et des relations entre ces termes qui la composent. Donc si le concept est une théorie alors il n'est pas possible d'expliquer son apprentissage comme on apprend un terme dans une théorie scientifique.

L'interprétation fonctionnelle du concept aux dépens de son application – de son extension – semble confirmée par la théorie de la catégorisation acceptée par la théorie des concepts. L'appel à des propriétés et des caractères abstraits – sous forme de propriétés cachées – indique que la description passe par la fonction caractéristique à laquelle l'objet se conforme. Ce qui importe donc c'est la manière d'expliquer l'objet – qui à lui accoler des propriétés qu'il ne manifeste en fait pas – plutôt que d'essayer de constituer des descripteurs à partir de ce que manifeste l'objet. En somme l'explication – voire même *a priori* prime sur l'expérience [Margolis and Laurence, 1999, p. 45]. C'est ce type même d'explication qui permet à un zoo de vendre une mule peinte pour un zèbre.

Mais les agents cognitifs évoluent, l'expérience leur fait comprendre certaines choses qu'ils ignoraient lorsqu'ils étaient enfants. Si les enfants peuvent accepter que les mules peintes soient des zèbres, les adultes savent qu'il y a des zèbres et qu'il y a des mules dans le monde, et que quelque fois les directeurs de zoo ou bien pour prendre les visiteurs pour des têtes de mules ou bien à cause de restrictions budgétaires, mettent des mules peintes à la place des zèbres. Les concepts des enfants ne sont pas les mêmes que les concepts des adultes. La théorie des concepts explique cela par la différence de théories possédées par les adultes et celles possédées par

les enfants. Celles des enfants étant naturellement plus pauvres et moins riches que celles de leurs aînés. Les théories des adultes s'enrichissent et se développent à partir de l'expérience – ce qui suppose que l'extension joue un rôle dans la constitution des théories, sur le modèle empirique, et donc que la curiosité pousse les adultes à expliquer d'avantage de phénomènes que les enfants – et également à partir d'autres concepts et théories. Cela explique en partie la compositionnalité des concepts : certains concepts combinés entre eux au sein d'une théorie forment un autre concept. Mais cela est plus un *usage* de la compositionnalité qu'une explication de celle-ci. D'autre part si les concepts sont des théories et non pas des termes au sein d'une théorie, comment expliquer la compositionnalité de deux systèmes déductifs différents ? Il semble que le problème de l'incommensurabilité des théories empêche toute composition et mise en commun de celles-ci.

Le principal problème de la théorie des concepts comme théorie est que l'on ne sait pas très bien ce qu'est une théorie. Est-ce que le concept est un terme qui apparaît dans une théorie comme le concept MAMMIFÈRE apparaît dans une théorie biologique, ou bien est-ce que le concept est une théorie en bon est dû forme avec pour chaque concept un jeu de règles de déduction, d'axiomes et de théorèmes. Si le concept est un constituant d'une théorie alors il n'est pas une théorie, s'il est une théorie à part entière, qu'est-ce qui le compose ?

De plus, si le concept est une théorie, alors pour maîtriser le concept il faut que l'agent cognitif puisse rendre compte de cette théorie, c'est-à-dire en donner les règles de déduction, les axiomes et les théorèmes. Or, même lorsqu'il s'agit de théories naïves, il se trouve que l'agent cognitif n'est pas nécessairement en mesure d'explicitier l'usage et l'application qu'il fait du concept. Une théorie du concept ne semble pas exiger un principe épistémique fort, mais simplement le principe épistémique faible. Un agent peut utiliser un ordinateur, posséder le concept ORDINATEUR sans être en mesure d'expliquer comment un ordinateur marche.

Mais le plus gênant pour la théorie des concepts comme théorie est l'explication de l'erreur par un agent dans l'application ou l'usage d'un concept. Si le concept est une théorie, alors l'agent qui possède le concept α possède de fait la théorie α . Or si α est une théorie elle doit se conformer aux principes de constitution d'une théorie, entre autre être consistante. Et si α est

consistant alors l'agent peut l'utiliser sans problème. Et si ce concept est étrange ou s'il est faux, si par exemple il dit que les baleines sont des poissons, l'agent est en droit d'utiliser sa propre théorie. Mais s'il est possible de dire qu'un agent utilise mal un certain concept, cela signifie qu'il est possible d'évaluer et de juger un concept. Or pour comparer deux théories il faut ou bien une troisième théorie ou bien une norme, mais comment comparer deux concepts? Qui les comparent? L'agent peut-il avoir accès à son concept et à ce que son concept est ou devrait être pour un autre agent? Si le concept est une théorie il est toujours possible à l'agent qui le possède de rajouter *a posteriori* une hypothèse, un axiome ou une règle *had hoc* qui justifie ou rétablit l'usage de son concept. De fait il n'y a pas de mauvais concept ou de concept faux. Il y a peut-être des concepts meilleurs que d'autres, mais la pertinence est différente de la vérité, et de cela la théorie des concepts comme théorie ne peut en parler.

3.3 Exemples

Théorie des Exemples : *le concept est n'importe quel élément d'un ensemble qui permet de représenter celui-ci.*

Le principal problème de la théorie classique et de la théorie des concepts comme théorie est qu'elle se soucie beaucoup de la description des éléments mais assez peu des éléments décrits. Or si l'agent fait ou utilise des descriptions, des concepts pour représenter l'environnement c'est parce qu'il est en relation avec ces éléments. Fortes de ce constat, certaines théories du concept préconisent une approche extensionnelle du concept, et proposent de constituer celui-ci non pas à partir des descripteurs mais des éléments décrits. En somme elles utilisent une méthode « bottom up » plutôt que « top down ».

L'idée de la *théorie des exemples*, proposée par [Medin and Schaffer \[1978\]](#) est de fonder le concept non plus sur les propriétés qu'un objet doit manifester pour appartenir à un concept, mais de se baser sur les instances du concept dont l'agent a fait l'expérience. Par exemple le concept CHIEN ne correspond plus à une définition ou à une théorie mais à l'ensemble des chiens que l'agent a rencontré et dont il se souvient. Aucune abstraction n'est faite sur ces occurrences du concept, si bien que toutes valent à part

égale pour représenter le concept. Lorsque l'agent lit dans un roman que le chien à mordu la petite fille, sans autre caractérisation du chien, il peut penser que c'est un caniche ou un terre-neuve cela ne change pas grand-chose puisque ce sont tous deux des chiens.

La théorie des exemplaies est une théorie empirique du concept, dans la lignée de celle de Locke, de Berkeley ou de Hume. Les concepts y sont conçus comme des représentations mentales – dont le paradigme est visuel, ce sont des images mentales – élaborées à partir de l'expérience sensible. Lorsque l'agent voit un chien, il en garde une trace dans son esprit, trace qu'il peut réactiver par la suite grâce à la mémoire. La dépendance du concept vis-à-vis de son extension – de ses occurrences – explique sa possession et sa genèse. Un agent aura le concept de ce dont il aura fait expérience, le concept provient donc de l'expérience et s'acquiert par elle. Un agent qui aura vu un caniche possédera le concept CHIEN dont sa représentation mentale sera un caniche, alors qu'un agent ayant fait l'expérience de xoloitzcuintles possédera également le concept CHIEN mais sa représentation mentale sera celle d'un xoloitzcuintle. Cela explique le contenu cognitif de l'agent, et comment deux agents peuvent posséder un même concept tout en lui associant un contenu différent. Cette différence d'expérience d'occurrence d'un même concept par des agents cognitifs différents explique également l'*effet de typicité*. Plus un élément sera proche des instances d'un concept déjà perçues par un agent, plus celui-ci reconnaîtra cet élément comme étant une occurrence de ce concept. L'environnement joue donc un rôle important dans l'apprentissage et le maniement des concepts. Dans un environnement où les caniches seront nombreux et les xoloitzcuintles rares, un agent aura une représentation mentale correspondante au concept CHIEN qui ressemblera plus à un caniche qu'à un xoloitzcuintle.

La catégorisation – le fait de subsumer un élément sous un concept – dépend donc des instances déjà présentes dans l'extension du concept. C'est en comparant l'élément à catégoriser avec les éléments que l'agent a en mémoire. Comme tous les exemplaies valent de manière égale pour représenter le concept, l'agent doit comparer l'élément nouveau à l'ensemble des éléments en mémoire. C'est-à-dire que si l'agent a vu 40 000 chiens dans sa vie et que devant lui se tient un teckel à poils durs, et bien il passera en revue les 40 000 occurrences de chiens, plus toutes celles de ce que n'est pas le teckel, c'est-à-dire, les chats, les poules, les veaux et les chars d'assaut,

jusqu'à ce qu'il en déduise que ce teckel ressemble plus à un caniche qu'à une motoneige ou que d'un volcan arctique.

La catégorisation passe donc par la comparaison et le degré de similitude entre l'élément à classer et les éléments déjà classés, en utilisant le *principe contrastif* introduit précédemment (2.5). Chaque élément perçu dans l'expérience est décrit par des descripteurs pondérés, l'ensemble des points d'équilibres relatifs qui émergent lors de la mise en rapport de tous les éléments déterminent des concepts, et tout nouvel élément, du fait de sa pondération se retrouve plus ou moins distant d'autres éléments dans l'espace de pondération. L'élément est considéré comme faisant parti du même concept que les éléments dont il est le plus proche.

Comme la comparaison implique l'ensemble des éléments présents de manière particulière dans l'espace cognitif de l'agent, aucun trait n'est perdu, ce qui permet de comparer de manière très fine les éléments entre eux. Ainsi un xoloitzcuintle ne paraîtra pas un chien étrange, mais un chien tel qu'il est peu représenté dans un environnement plein de caniches. Évidemment le problème est qu'il se peut que le xoloitzcuintle soit considéré par l'agent comme étant plus similaire à un chat qu'à un caniche, et qu'il soit mal catégorisé. Mais si la catégorisation ne dépend pas des caractères en tant qu'ils doivent décrire un élément mais en tant qu'ils sont manifestés par cet élément, alors si l'agent pense que xoloitzcuintle ressemble plus à un chat qu'à un chien, alors le xoloitzcuintle, pour cet agent, relèvera du concept CHAT et non de celui de CHIEN. La théorie des exemplaires considère donc que le concept n'est pas abstrait – sinon à titre théorique – mais qu'il se réduit à l'ensemble de ses membres, de fait aucun processus d'abstraction n'a lieu pour cette théorie en ce qui concerne le concept. L'inconvénient évidemment est que s'il faut passer en revue l'ensemble des éléments issus de l'expérience, cela peut prendre un temps conséquent à chaque fois, tellement conséquent même qu'il devient tout simplement impossible en pratique.

Le problème se pose également de savoir comment sera classé un élément d'un concept non déjà possédé par l'agent. Cet élément sera toujours près d'un élément que d'un autre, donc d'un concept. Par exemple l'ornithorynque peut être considéré comme un oiseau ou comme un mammifère, mais rien n'indique qu'il faille le classer à part dans un nouveau concept. Et

de ce fait aucun nouveau concept ne peut apparaître ou être créé. Se pose alors la question de savoir d'où proviennent les concepts utilisés par un agent cognitif. Sont-ils innés, proviennent-ils de l'évolution de la cognition ? Mais alors cela signifie qu'ils seraient indépendants de leur extension. Même si l'agent cognitif humain possédait le concept α du fait de l'évolution et de la proximité entre l'espèce humaine et de α la sélection du concept α dépend d'interactions les éléments manifestant α et l'espèce, et donc sur les caractères de ces éléments. De sorte que si un agent cognitif humain vient à être en relation avec un élément manifestant α alors il aurait de manière réflexe et innée telle et telle réaction prévisible. Mais l'agent possède cette *disposition* combien même il ne serait jamais en relation avec un quelconque élément manifestant α . Or cela n'est pas compatible avec la version de l'exemplarisme qui affirme que la catégorisation est basée sur la similitude avec l'expérience passée. À moins que des caractères soient extraits de chaque élément pour permettre de le décrire et de le comparer à d'autres. Mais alors ce n'est plus d'exemplaies dont il est question, mais de fonctions caractéristiques de ceux-ci, c'est-à-dire de caractères, et alors nous retombons sur une théorie classique ou une théorie théorie du concept. L'apprentissage artificiel ne marche que si au moins des seuils de pondérations sont donnés [Cornuéjols and Miclet, 2003] – c'est-à-dire qu'une limite est donnée au moins dans la granularité de la détermination des points d'équilibre, sans quoi le résultat final est que chaque élément devient son propre point d'équilibre, écueil du « plus pur nominalisme » [Hacking, 1993] ou bien au contraire il n'y a qu'un seul point d'équilibre, c'est-à-dire un seul concept. Dans les deux cas la classification est inutilisable et donc vaine. En fait, ce qui manque dans cette approche est l'*usage* du concept, de savoir pourquoi un agent cognitif peut avoir besoin de concept. La réponse à cette question implique de se pencher sur la *pragmatique* du concept, sa finesse et sa structure en dépendent.

Cependant, et c'est là la grande force de la théorie des exemplaies, la détermination des seuils, de la taille des ensembles et même des caractères choisis peut être laissé à la charge de l'agent cognitif. Un agent peut constituer les concepts dont il a besoin. Cela explique en partie le *contenu cognitif* du concept. L'autre aspect du contenu cognitif étant déterminé par l'expérience particulière de l'agent. Deux agents peuvent donc posséder le

même concept tout en y associant des contenus ou des usages différents. Bien que si le concept dépend de l'expérience et des occurrences perçues par l'agent, cela explique l'extension du concept, mais difficilement son usage. Comment l'agent peut-il penser que le chien qui vient de voir est un mammifère ? Il peut appliquer deux concepts différents à un même objet, ce qui est un chien est aussi un mammifère, mais cela n'explique pas le lien qu'il y a entre le concept CHIEN et le concept MAMMIFÈRE. Il y a une différence entre cette relation et celle qu'il peut y avoir entre le chien que l'agent vient de voir marquer son territoire contre son vélo et le fait que ce chien appartienne à la voisine qui vient quelque fois fumer sous la fenêtre. La première relation est nécessaire alors que l'autre ne l'est pas.

Si le concept dépend de l'expérience de l'agent, cela permet de comprendre pourquoi le concept est limité dans sa généralisation du point de vue cognitif. Un agent peut parfaitement utiliser et appliquer comme il faut un concept et puis soudainement, allez savoir pourquoi, considérer que le concept ne l'applique pas pour l'une de ses instances. Un agent peut dire ce qu'est un chien par exemple et appliquer le concept CHIEN dans la plupart des cas, sauf pour les xoloitzcuintles, qui sont pourtant des sujets canins. Il se peut par exemple que l'agent considère en fait que les chiens sont couverts de poils, et tout ce qui ressemble à un chien mais qui n'est pas couvert de poils n'est pas un chien, mais autre chose. De fait, même si l'agent est d'accord pour dire que le xoloitzcuintles ressemble à un chien, ce n'en est pas un, alors que le loup, qui n'en est pas un, en est un. La limitation de la généralisation vient alors du contenu de l'expérience de l'agent. L'agent a perçu plus de caniches que de xoloitzcuintles, et en a déduit que les chiens avaient des poils. Cependant, cette généralisation ne peut faire partie du concept ou de son contenu, puisque cela supposerait une *abstraction* sur les instances du concept, en particulier en extrayant des caractères de ces instances. Mais la théorie des exemplaires ne peut faire cela, ou du moins pas au niveau du concept, puis que cela signifierait que le concept n'est pas un ou plusieurs *exemplaires* mais certaines propriétés de ceux-ci, par exemple comme celles des exemplaires situés aux points d'équilibre de l'espace pondéré, comme le défendra la théorie des prototypes.

L'avantage de la théorie des exemplaires est qu'en acceptant la limitation de la généralisation du concept, elle rend compte d'un phénomène psychologique : le fait que les agents ne respectent pas scrupuleusement

les inférences logiques formelles, même s'ils sont prêts à les reconnaître comme des règles valides et qu'il est utile de suivre. Cette limitation de la généralisation, due au fait qu'il est même difficile de parler de généralisation dans le cadre de la théorie des exemplaires du fait de l'absence de théorie de l'abstraction, fait toute la particularité de cette théorie. La catégorisation, entendue ici comme l'assignation d'un élément à un ensemble, se faisant par comparaison de cet élément avec l'ensemble des éléments déjà perçu, se fait sur l'ensemble des caractères de l'ensemble des éléments, de fait, contrairement à une théorie de l'abstraction [Fine, 2002], aucune information n'est perdue. L'assignation est donc très fine. Mais il n'est plus certain alors que l'on peut parler d'assignation, sinon à titre métaphorique.

Si l'agent compare l'élément à classer avec l'ensemble des éléments dont il dispose déjà, alors l'apprentissage du concept devient clair. Un agent acquiert un concept par l'expérience. Le concept n'étant alors qu'un ensemble d'éléments rassemblés du fait de leur similitude, d'après les données de l'expérience, c'est-à-dire les propriétés saillantes du percept. La théorie des exemplaires est donc une théorie extensionnelle du concept dans la lignée de l'empirisme. Mais comme la théorie des exemplaires se refuse à toute abstraction, c'est-à-dire à tout traitement du contenu de l'expérience, cela signifie que le concept ne requiert aucun engagement épistémique, ni fort ni faible, de la part de l'agent. La comparaison peut se faire sur la simple base du percept. Par conséquent, la théorie des exemplaires ne dresse pas de limite et de différence, ni de genre ni de nature, entre le concept et le percept, entre le contenu conceptuel et le contenu non-conceptuel. La théorie des exemplaires n'est qu'une théorie des protojugements au niveau du percept, c'est-à-dire de la proto-catégorisation au niveau non-conceptuel, sans recours à la représentation. Elle n'est pas encore une théorie du concept.

3.4 Prototypes

Théorie des Prototypes : *le concept est l'ensemble moyen des caractères possédés par les membres d'un ensemble.*

La théorie des prototypes, développée par Rosch [1975] et Hampton [1982], est un raffinement de la théorie des exemplaires. Les concepts ne sont

plus représentés par l'ensemble de leurs occurrences, mais par un *prototype*, c'est-à-dire un élément qui a les traits les plus partagés par les occurrences du concept. Dans la théorie des exemplaires, un élément appartenait au même concept que les éléments dont il était le plus similaire. Mais la théorie des exemplaires se refusait à toute abstraction sur ce concept. La théorie des prototypes affirme simplement que le calcul de la similitude entre les éléments appartenant à un même concept, forme un point d'équilibre [Nash, 1950] sur la trame des descripteurs pondérés de l'ensemble des caractères des éléments catégorisés. Si l'on s'accorde un principe d'abstraction, à partir de ce point d'équilibre, alors il est possible de former un élément – qui n'est pas nécessairement présent dans l'ensemble, et c'est en ce sens qu'il est un abstrait – qui a les caractères moyens de l'ensemble des éléments regroupés dans un concept. L'appartenance au concept n'est plus déterminée par une liste stricte de caractères nécessaires et suffisants, comme le demandait la théorie classique et comme le demande une théorie classique des ensembles [Krivine, 1998], de fait la théorie des prototypes se base plutôt une théorie des sous-ensembles flous [Zadeh, 1965]. Ainsi un élément peut appartenir plus ou moins à un ensemble suivant la proximité qu'il entretient avec le prototype. Par exemple, un dodo et un moineau sont tous les deux des oiseaux, mais le moineau sera considéré, par un agent cognitif comme étant « plus » un oiseau que le dodo, parce que le moineau ressemble plus au prototype d'oiseau, qui est, sous nos latitudes pas très éloigné du rouge-gorge. Cette reconnaissance cognitive d'appartenance d'un élément à un concept est calculée suivant le temps de réaction de l'agent dans une tâche de classification. L'agent mettra plus de temps à penser que le dodo est un oiseau qu'il ne le fera pour le moineau. Cela montre que l'agent est surpris et a besoin de considérer les caractères du dodo avant de le catégoriser, alors que la routine de classification pour le moineau est plus rapide, donc les caractères du moineau surprennent moins l'agent que ceux du dodo, c'est le fameux *effet de typicité*.

L'appartenance de l'élément au concept n'étant pas ensembliste mais se faisant sur la base du degré de ressemblance entre l'élément et le prototype selon l'agent, les propriétés des concepts – c'est-à-dire des prototypes – ne sont pas nécessairement transitives dans le cas du recoupement de concepts. Par exemple, si un élément de l'environnement est dit être un siège et que les sièges sont dit être des meubles, alors il s'en suit par transitivité

de propriétés que l'élément en question est un meuble (cf. 3.1). Un agent pourra ainsi penser qu'un siège de voiture est un siège sans penser que c'est un meuble.

Toutes les instances d'un concept n'ont pas la même valeur au sein de ce concept. Plus l'élément est proche du prototype plus il sera considéré appartenir au concept. Cela explique l'effet de typicité mais explique également que dans le cas d'une révision ou d'un changement de concept, certains éléments sont sacrifiés plus rapidement que d'autres. Les éléments les plus proches du prototype seront préservés plus facilement que ceux qui en sont plus éloignés. Un agent sera plus enclin à reconnaître qu'un tapis roulant n'est pas un véhicule qu'un vélo n'est pas un véhicule.

Le prototype est le meilleur élément d'un concept, celui qui se trouve au point d'équilibre des caractères des occurrences du concept. Le prototype change donc à l'arrivée de chaque nouvel élément dans le concept, mais comme il est situé au point d'équilibre il change relativement peu, à moins que le concept lui-même soit amené à changer de manière radicale. La notion de prototype s'inspire de la notion d'« *air de famille* » de Wittgenstein [1958] et de la notion de *stéréotype* de Putnam [1990].

Le prototype est constitué à partir du calcul de degré de similitude de l'élément par rapport au poids moyen de ses caractères :

$$\text{(Similitude)} \quad D(i, j) = [\sum_{k=1}^n |X_{ik} - X_{jk}|^r]^{1/r} \quad (3.4)$$

où $D(i, j)$ est la distance entre les éléments i et j , n est le nombre de dimension de l'espace pondéré des traits caractéristiques, X_{ik} et X_{jk} sont respectivement les valeurs pour i et j du nombre r de leurs caractères respectifs dans l'espace de dimension n .

À partir de l'ensemble des distances, le point d'équilibre est calculé suivant l'indépendance de la fonction de calcul de l'appartenance d'un élément à un ensemble selon la similitude :

$$\text{(Point d'équilibre)} \quad \tilde{x} := \frac{dx}{dt} = f(t, x) \quad (3.5)$$

si $f(t, x) = 0$ alors cela signifie que la fonction f est indépendante vis-à-vis du paramètre t , et cela signifie que x équilibre la fonction f . On note alors \tilde{x} ce point qui est le *point d'équilibre* de la fonction f . Ce point peut n'être

instancié par aucun élément présent dans le concept, mais il sert de norme pour déterminer la similitude entre les éléments. Dans l'espace pondéré des éléments plusieurs points d'équilibres relatifs émergeront, suivant le nombre de dimensions – le nombre de caractères – de l'espace. Le nombre de caractère déterminant la finesse de granularité de la description des éléments de l'environnement. Ainsi à chaque point d'équilibre correspond un concept. Tous les éléments présents dans l'espace pondéré sont décrits d'une manière ou d'une autre, même les cas particuliers situés à équidistance de deux points d'équilibre par exemple, du fait même que le calcul de similitude ne relève pas d'une théorie des ensembles – qui supposerait une appartenance stricte – mais d'une appartenance relative. Un élément situé à équidistance de deux points d'équilibre sera dit appartenir au deux – quand bien même ils seraient mutuellement contradictoires – ou bien arbitrairement à un seul.

La théorie des prototypes va donc plus loin que la théorie des exemplaires en caractérisant le point d'équilibre et en en faisant le représentant du concept. De sorte que le calcul de la similitude se fait ensuite non plus en fonction de la comparaison de l'ensemble des éléments entre eux mais de chaque élément par rapport aux points d'équilibres.

$$D(i, \tilde{x}) = \left[\sum_{k=1}^n |X_{ik} - X_{\tilde{x}k}|^r \right]^{1/r} \quad (3.6)$$

Le calcul de l'appartenance d'un élément à un concept doit se faire par la comparaison de cet élément à l'ensemble des points d'équilibres présent dans l'espace pondéré, c'est-à-dire en tenant compte de tous les concepts disponibles à l'agent. De sorte que les concepts sont dépendants les uns les autres, au moins quand à la détermination de leur extension. Cela implique une forme de holisme.

La théorie des prototypes parvient à rendre compte et à expliquer comment les agents cognitifs catégorisent effectivement l'environnement, notamment grâce au principe de similitude et à la constitution d'éléments prédominants servant de référence au sein d'une catégorie : le prototype. La mise en évidence du prototype et de son utilisation dans l'assignation d'un concept explique l'effet de typicité et les différences de réaction d'un même agent face à différences instances d'un même concept. Contrairement

aux prédictions de la théorie classique, le concept n'est pas uniforme et ses instances n'ont pas toutes le même poids d'un point de vue cognitif. En mettant l'accent sur ce point, la théorie des prototypes n'essaye pas de faire une théorie de la catégorisation d'un point de vue logique – comme le fait la théorie classique en adoptant une théorie classique des ensembles – mais à expliquer la classification d'un point de vue psychologique.

Le prototype dépend des éléments déjà catégorisés par l'agent, puis qu'il en est la synthèse. Cela signifie que le prototype dépend de l'expérience propre de chaque agent, un agent qui n'aura vu que des dodos aura un prototype pour le concept OISEAU plus proche du dodo qu'un agent qui n'aurait vu que des moineaux. La théorie des prototypes, se basant sur l'extension du concept plus que sur son intension, justifie la différence entre les contenus cognitifs de différents agents pour un même concept du fait de la différence d'expérience et d'histoire personnelle de ces agents.

La fonction du concept est, selon la théorie des prototypes, essentiellement de permettre de catégoriser les éléments de l'environnement. Nous avons vu que c'était là la faiblesse de la théorie des exemplaires, qui parlait de classification sans se donner aucun principe d'abstraction. La théorie des prototypes, en admettant l'abstraction sur le point d'équilibre des caractères moyens partagés par les éléments d'un ensemble, peut constituer une véritable classification avec une hiérarchie de concept. D'autre part, en partant de l'extension du concept à partir de ses occurrences dont l'agent fait l'expérience, la théorie des prototypes rend compte du premier niveau de la classification, celui des objets classés. La base de la classification est constituée des éléments issus de l'expérience, et il est possible de justifier cela en recourant au contenu non-conceptuel et du traitement de l'expérience par la sensation. Cette interprétation est justifiée par le fait que les éléments seront anthroposcopiques (à échelle humaine), et traiter selon les théories naïves. La hiérarchie des concepts se fait sur la base des prototypes. À partir de plusieurs prototypes, grâce au principe d'abstraction il est possible de former des prototypes de prototypes, c'est-à-dire des concepts de concepts. C'est ainsi que les oiseaux et les mammifères pourront être des animaux. La théorie des prototypes prédit que le niveau le plus commun et le plus facile à acquérir dans la classification sera celui des prototypes de niveau

moyen. Par exemple il sera plus facile pour un agent cognitif d'acquérir le concept OISEAU que le concept MÉSANGE À TÊTE NOIRE qui est trop précis et donc demande une sélection plus fine de caractères, ou que le concept ÊTRE VIVANT qui est trop général et fait abstraction de certains caractères généralement retenus par l'agent. La théorie des exemplaires prédisait que les éléments particuliers seraient acquis en premier, et la théorie classique affirmait que les concepts de haut niveau seraient plus simples à apprendre. Or il semble que les expériences psychologiques donnent raison à la théorie des prototypes.

La simple expérience d'élément de l'environnement n'est pas suffisante pour constituer un terme général au sens intensionnel du terme – un abstrait – mais est suffisante pour constituer un prototype [Sylvand, 2000, 2001]. C'est-à-dire que le prototype sera dépendant des instances à partir desquelles il est constitué, alors que l'abstrait ne dépend d'aucune occurrence mais les détermine. Cela est suffisant au moins pour rendre compte des théories naïves et expliquer les limites cognitives de la généralisation.

Cette généralisation basée sur le calcul de distance et de degré de similitude entre un élément et un référent, est relativement simple à mettre en œuvre et peut être automatisée dans un réseau de neurones par exemple [Jodouin, 1994], [Cornujéols and Miclet, 2003], l'implémentation dans un système cognitif semble donc plausible et nettement moins coûteuse que le procédé par comparaison de l'élément à catégoriser avec l'ensemble des éléments en mémoire, comme le veut la théorie des exemplaires. D'autant plus plausible est cette implémentation, que le principe de calcul de similitude semble être utilisé au niveau non-conceptuel dans le traitement des données sensorielles. Ce principe n'est donc pas dépendant du niveau conceptuel, et pourrait provenir d'un module logique cognitif indépendant [Fodor, 1974, Sperber, 2000]. Le principe de calcul de similarité entre les éléments au niveau conceptuel dépend des éléments perçus par l'agent, et donc est contextuel, ce qui explique le caractère borné des inférences et de la généralisation à partir du concept par un agent cognitif. Un agent peut posséder les concepts ÉTOILE DU MATIN et ÉTOILE DU SOIR, les appliquer et les utiliser correctement, sans nécessairement faire le lien entre les deux. Cela provient du fait qu'il est possible qu'un même élément de l'environnement supporte différents prototypes [Murphy, 2002], du fait de la non-exclusivité

des classes et que la théorie ensembliste utilisée n'est pas la théorie classique des ensembles mais plutôt une théorie proche de celle des sous-ensembles flous.

La détermination de la co-référentialité des concepts exige d'avoir recours au *contenu objectif* du concept, c'est-à-dire à sa référence indépendamment de ce qu'elle est pour et selon l'agent qui possède le concept. Ce contenu n'est pas et ne peut pas être donné par le prototype du fait que celui-ci est déterminé par les occurrences perçues par l'agent. Le prototype est dépendant du mode de présentation de l'élément, et donc relatif à l'expérience de l'agent, il ne peut être objectif. L'agent ne peut déterminer la co-référence de deux concepts que si un mode de présentation commun – selon lui – aux deux concepts, vient à surgir. Cela implique également qu'aucun concept vide ne peut être constitué par l'agent, ni aucun concept abstrait, c'est-à-dire qui n'est pas raccordé d'une manière ou d'une autre à l'expérience sensible. Or certains concepts tels que les concepts mathématiques ou logiques par exemple ne semblent être directement expérimentables.

La théorie des prototypes peut se défendre en affirmant qu'il est possible de posséder un concept – et donc un prototype – d'un élément dont l'agent n'a pas fait l'expérience. Un agent peut posséder le concept LICORNE ou MONTAGNE D'OR en combinant des prototypes qu'ils possèdent déjà, entre deux. Après tout un cheval auquel on a collé une corne de narval est une licorne. Mais si cela semble se comprendre pour des abstraits « sensibles » [Hume, 1983], les abstraits au sens propre du terme comme les abstraits mathématiques sont rebelles à une réduction de ce type [Hilbert and Bernays, 2001a, Fine, 2002].

Pour composer des concepts, par exemple le concept CHEVAL et le concept CORNE DE NARVAL pour obtenir le concept LICORNE, suppose que les concepts se composent. En comme les concepts sont, selon la théorie des prototypes, des prototypes, cela suppose que les prototypes se composent entre eux. Des expériences psychologiques [Prinz, 2002, pp. 39-40] montrent que la théorie des prototypes rend compte de phénomènes assez étranges de non monoticité dans la composition des concepts. Le concept complexe OUTIL QUI EST AUSSI UNE ARME (TOOL THAT IS ALSO A WEAPON) indique de manière analytique que tout ce qui tombe sous ce concept est un outil qui est par ailleurs une arme. Mais l'expérience montre que les agents cognitifs ne le

voient pas de cet œil-là. Ils sont prêts à reconnaître par exemple qu'un tournevis est un outil – qui sert principalement à tourner les vis – et qu'il peut parfois être utilisé pour trucider les vieilles dames, et donc servir d'arme. De fait, les agents reconnaissent qu'un tournevis est un outil qui est aussi une arme, et assigne ce concept à cet objet. Par contre lorsqu'on leur demande si un tournevis est une arme, la plupart des agents qui reconnaissaient que le tournevis était un outil qui était aussi une arme, répondent que non, un tournevis n'est pas une arme. Or de manière logique, l'affaiblissement de la conjonction devrait préserver sa valeur.

$$\begin{array}{l|l}
 1 & \text{tournevis} = \text{outil} \wedge \text{arme} \\
 \hline
 2 & \text{tournevis} = \text{outil} \\
 3 & \text{tournevis} = \text{arme}
 \end{array} \tag{3.7}$$

puisque selon la table de la conjonction, la conjonction est vraie si ses deux membres sont simultanément vrais :

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	0

TAB. 3.1 – Table de vérité de la conjonction

D'un point de vue logique, si un élément est décrit par la conjonction de p et de q cela implique que l'élément en question est à la fois p et q , et donc il est p en particulier. Cependant, la cognition n'est pas la logique formelle. Le contexte et la pertinence de l'information sont pris en compte par l'agent, alors qu'elle ne l'est pas par le calcul logique. L'agent juge que si l'objet est décrit par une conjonction c'est qu'il doit l'être ainsi, et que si on le décrit avec un seul de ses caractères alors on ne le décrit pas suffisamment.

La théorie des prototypes arrive à expliquer ce phénomène. L'appartenance d'un élément à un concept est déterminée par le degré de similitude qu'entretient cet élément avec le prototype du concept. Il y a comme un seuil d'entrée dans l'extension d'un concept, ce seuil est relatif à la distance entre deux concepts. Pour établir ce seuil, on détermine le poids des descripteurs

de chaque concept et le degré nécessaire de ressemblance pour assigner ce concept à un objet. Ce seuil a pour limite l'équidistance entre deux concepts, c'est-à-dire qu'un objet doit avoir au moins 50% de la valeur de poids de description pour entrer dans un concept comparé à d'autres concepts. 50% au moins parce qu'un élément ne possédant que 50% des traits d'un concept et 50% des traits d'un concept peut appartenir au deux, puis que nous avons vu que les classes n'étaient pas nécessairement exclusives. D'autre part si un élément ne partageait que 0,1% des poids de description d'un concept mais 0% des poids de tous les autres concepts, alors ce concept partagerait 100% des traits du concept en question relativement aux autres, et relèveraient donc de ce concept.

Sur cette base, imaginons que le seuil d'entrée dans le concept OUTIL soit conditionné par les traits t_1, t_2, t_3 et que celui du concept ARME par les traits t_4, t_5, t_6 . Imaginons maintenant que le seuil d'entrée pour le concept OUTIL QUI EST AUSSI UNE ARME est conditionné par les caractères t_1, t_2, t_3, t_5 . Supposons par ailleurs que les traits t_1, t_2, t_3 soient nécessaires et suffisants pour déterminer l'appartenance au concept OUTIL mais que le trait t_5 ne soit que nécessaire mais non suffisant pour déterminer l'appartenance au concept ARME. Le tournevis peut satisfaire les traits t_1, t_2, t_3, t_5 ce qui fait de lui un outil qui est aussi une arme, mais pas une arme [Prinz, 2002]. Mais cette justification est valable si et seulement si l'on admet que les traits ne déterminent pas de manière formelle l'appartenance au concept, car d'un point de vue logique, si le trait t_5 n'est pas suffisant pour permettre d'assigner le concept ARME à un élément, cela signifie que l'ensemble des traits t_1, t_2, t_3, t_5 ne l'est pas plus pour permettre cette assignation, et donc le concept OUTIL QUI EST AUSSI UNE ARME est nécessaire et suffisant pour désigner un outil, mais pas une arme.

Le problème est que la constitution du prototype semble inutilement compliquée. En effet il suppose que l'agent passe en revue l'ensemble des éléments qui apparaissent dans l'extension du concept, qu'il les décrive et assigne un poids pour chacun de leur trait, qu'ensuite il fasse la moyenne de tous ces poids, utilise le principe d'abstraction, et constitue le prototype, après quoi il applique ce prototype à chacun de ces éléments et en évalue la distance, et en fonction de la pondération des poids moyens, évalue chaque trait de chaque élément. Et comme le prototype dépend des occurrences du

concept – puisque ce n'est pas véritablement un abstrait – il faut recommencer cette opération à chaque usage du prototype. En somme cette technique n'est pas moins onéreuse que la théorie des exemplaires. De plus la théorie des prototypes fait appel à deux types de calculs : l'un pour constituer le prototype, l'autre pour établir de degré de similitude entre l'élément et le prototype. Pourquoi alors ne pas directement comparer l'élément à l'extension du concept comme le préconise la théorie des exemplaires ?

Ensuite le statut du prototype n'est pas très clair. Il est un faux abstrait dans la mesure où il dépend des éléments perçus par l'agent, et aucune généralisation inductive ou même abductive ne détermine un terme indépendant extensionnellement. Un abstrait ne peut être conditionné par l'expérience. Cependant le prototype n'est pas non plus strictement réductible à l'expérience et à son donné, puisqu'il suppose une synthèse des caractères de ce contenu. Toute l'ambiguïté du problème est là : le prototype prétend se baser sur le contenu de l'expérience et pourtant il fait appel à un espace pondéré de caractères. Mais en établissant la moyenne des traits les plus représentés, le prototype affaiblit la représentation du concept en délaissant les caractères les moins pondérés, sans établir des conditions nécessaires et suffisantes pour établir l'appartenance au concept. Les traits retenus par le prototype ne sont pas nécessairement pertinents à moins qu'il y ait des contraintes sur l'espace pondéré lui-même, mais c'est là une condition supposée par le prototype et non une chose à laquelle il peut pourvoir. Le fait de mesurer quarante centimètres au garrot peut être un trait du prototype CHIEN, mais ce n'est pas un trait très pertinent. D'autre part le prototype changeant suivant le donné de l'expérience, ses modulations peuvent donner des conjonctions de caractères étranges, comme le couteau sans lame auquel il manque le manche ou bien l'homicide sans cadavre, arme ni mobile. Genre de description pas très efficace.

L'avantage de la théorie des prototypes est qu'elle ne suppose pas que le concept soit clairement et systématiquement défini, et c'est ce qui lui permet de rendre compte de l'effet de typicité ou de la non-transitivité des propriétés. Cependant cela ne doit pas nécessairement impliquer que le concept est indéterminé. L'effet de typicité indique simplement que pour un agent cognitif une occurrence du concept est plus représentative du concept qu'une autre. Mais cela marche même pour les concepts clairement définis, tels que les artefacts par exemple. Une théorie inspirée de la théorie

classique, qui imposerait des traits nécessaires et suffisants pourrait très bien expliquer ce phénomène de typicité en distinguant les caractères de leurs instanciations. Il semble donc plus que l'effet de typicité dépend de l'*application* du concept que de la constitution du prototype, et donc que c'est le contenu cognitif qui peut en rendre compte et non pas le prototype.

La question centrale de ce travail est celle de l'explication de l'erreur et du changement de concept. Comment la théorie des prototypes les explique-t-elle ? L'erreur de concept consisterait à appliquer un concept à un élément qui n'en relèverait pas. Dire par exemple qu'une baleine est un poisson. Or selon la théorie des prototypes, appliquer un concept consiste à comparer l'élément avec un prototype. Ou bien l'élément possède un poids de description suffisant pour entrer dans l'extension du concept, ou bien non. Cela se faisant par comparaison, une baleine peut très bien ressembler à un poisson et donc relever, selon l'agent, du concept POISSON. Mais cela provient de l'objet baleine lui-même, pas de l'agent. En fait l'agent ne se trompe jamais. La seule erreur consiste à croire que le simple degré de similitude est suffisant pour attribuer un concept à un élément. La théorie des prototypes explique une routine, une réaction conditionnée d'un agent face à un objet du monde, mais pas une conceptualisation de celui-ci. La conceptualisation doit prendre en compte non seulement l'application du concept, mais également son usage. Et cela est particulièrement visible dans le cas des concepts abstraits, tels que les concepts mathématiques. Quel est le prototype du chiffre 2 ou du cardinal \aleph^0 ? Si les prototypes semblent être efficaces pour rendre compte des « concepts naïfs » elle ne semble pas l'être pour les concepts plus rigoureux ou formels.

La théorie des prototypes semble consciente de ses limites. Lorsque [Fodor \[1998, p. 94\]](#) l'accuse, de manière abrupte, de ne pas pouvoir rendre compte de la compositionnalité :

prototypes don't compose ⁴

[Hampton \[2004\]](#) répond en détaillant la stratégie de la théorie des prototypes, stratégie qui fait appel justement à l'espace pondéré de caractères.

Les prototypes ne composent pas au sens où le concept BOUTON DE MANCHETTE n'est pas composé des prototypes de BOUTON et de MANCHETTE,

⁴“Les prototypes ne se composent pas.”

puisque cela donnerait quelque chose comme un manche de chemise avec des boutons de rougeole ce qui ne correspond pas tout à fait à des boutons de manchette. La composition des concepts suppose une mise en commun des caractères compatibles des composants. L'élément désigné par le composé se situe à l'intersection des composants : il en partage les caractéristiques. Cela suppose donc que les caractères des composants soient dressés et comparés. C'est là une contrainte non pas sur le prototype mais sur l'espace pondéré de caractère sur lequel il s'appuie.

Hampton [2004] le reconnaît et suggère que la composition des concepts ne fasse pas seulement appel à l'extension – comme le fait la théorie des prototypes – mais également à l'intension du concept. Pour composer deux concepts il faut d'abord en abstraire les caractères, comparer les caractères des deux concepts, ne retenir que les caractères compatibles, les unir ensembles puis appliquer le terme composé de ces caractères à l'environnement pour en déterminer une extension. Cette approche est *a prioriste* et similaire à la théorie classique des concepts. Ce n'est qu'ensuite, lorsque ce terme a été appliqué et que des occurrences en sont reconnues qu'il est possible d'en constituer un prototype. La théorie proposée par Hampton, l'un des pères de la théorie des prototypes, n'est donc plus strictement une théorie des prototypes, mais une théorie mixte.

3.5 Empirisme conceptuel

Empirisme conceptuel : *tout concept provient de l'expérience.*

L'empirisme conceptuel est une théorie défendue par Prinz [2002] mais qui est fortement inspirée par l'empirisme anglais du dix-septième, notamment de Locke, de Berkeley et Hume. Il faut noter que les empiristes du dix-septième parlent d'*Idée* et non pas de concept, et ce n'est pas qu'une question de terminologie. Le problème auquel Locke entend apporter une réponse est celui de la querelle des universaux. La question du concept est une question d'architecture cognitive. Calquer le concept sur l'Idée serait considéré que le concept est simplement un terme général abstrait, en oubliant ses autres fonctions et caractéristiques, notamment le fait qu'il est un élément de l'espace cognitif de l'agent. La constitution de l'idée fait appel à des capacités de reconnaissance, d'individuation, d'identification

et d'abstraction que l'étude du concept cherche justement à expliquer. En somme l'Idée suppose le concept.

Prinz reprend l'idée de l'*image mentale* de l'empirisme et de la théorie des exemplaires :

When we introspect during thought, all we find are mental images, including auditory images of natural language sentences (subvocal speech). With no phenomenal traces of non-sensory-representations, it is tempting to conclude that all thought is couched in perceptual imagery ⁵. [Prinz, 2002, p. 73]

Cette notion d'image mentale est séduisante. Il est vrai que lorsqu'on pense à quelque chose on a l'impression de le voir ou de l'entendre dans sa tête. Du moins pour des objets que l'on peut d'habitude percevoir. J'avoue, pour ma part, ne pas avoir beaucoup d'images en tête lorsque je pense au théorème de Ramsey. Lorsque l'on parle d'« imagination » il ne paraît pas inopportun, mais cela n'implique pas que toute pensée, tout concept, fasse nécessairement appel à l'imagination. Prenez le cas d'un roman. Vous lisez les aventures de Miyamoto Musahi et l'on vous dit qu'il se bat avec un sabre en bois. Je sais à peu près à quoi ressemble un sabre de combat japonais du seizième siècle, je suis surpris de savoir qu'on utilise aussi des sabres en bois, qui me semblaient être des sabres d'entraînement, et non pas des sabres en acier damassé. Mais ce qui change avec cette nouvelle information n'est pas mon image, je n'ai pas changé la lame en acier damassé du sabre dans ma tête par une lame en bois, ce que je change c'est un réseau d'information qui me permet de relier ce sabre aux autres indications fournies. Sinon mon sabre dans la tête ressemblerait au couteau sans lame auquel il manque le manche, et je ne pourrais pas en faire grand-chose. Le fait qu'il soit difficile de couper quoi que ce soit avec un tel couteau n'est pas donné par l'image mentale, mais par la trame des inférences accessible depuis ce concept. Je n'ai pas plus d'image mentale de Musahi que je n'en ai du cardinal \aleph^0 . Bien sûr si l'on me demande de dessiner Musahi je pourrais représenter quelque chose, de même que je pourrais peut-être faire un schéma pour expliquer le cardinal \aleph^0 mais c'est là une autre question et une autre tâche. Mais

⁵“Lorsque nous introspectons quand nous pensons, tout ce que nous trouvons ce sont des images mentales, incluant des images auditives des phrases de la langue naturelle (langage subvocal). Sans traces phénoménales de représentations non sensorielles, il est tentant de conclure que toute pensée est couchée en imagerie perceptuelle.”

lorsque je lis le roman, je ne me soucie pas de la couleur ou du motif de son kimono tant qu'il n'est pas précisé, parce que cette information ne me manque pas nécessairement pour comprendre l'histoire.

L'empirisme conceptuel affirme que tout ce qui comprit dans le concept provient de l'expérience :

The Perceptual Priority Hypothesis : Nothing is in the intellect that is not first in the senses (*nihil est in intellectu quod non fuerit in sensu*) ⁶. [Prinz, 2002, p. 75].

Rien n'est dans l'intellect qui n'est d'abord dans les sens. Cet axiome empiriste est un refus catégorique de tout innéisme. Mais s'il peut se comprendre pour l'Idée et dans le débat sur les universaux [Sylvand, 2000], il l'est beaucoup moins pour les concepts. En effet le concept suppose – nous l'avons vu pour la théorie des prototypes par exemple – un processus d'abstraction. Et ce processus n'est pas donné par la sensation, mais permet de traiter le donné empirique. S'il est vrai que le contenu non-conceptuel se retrouve au niveau du concept, cela ne signifie pas pour tant que le concept est entièrement réductible au contenu non-conceptuel. Mais Prinz a la sagesse de rester flou. Il parle d'« *intellect* » et non de concept, et refuse par ailleurs toute notion de contenu non-conceptuel.

“Nonconceptual” implies a too sharp distinction between perceptual and the conceptual ⁷. [Prinz, 2002, p. 112].

Mais c'est là une conception erronée du contenu non-conceptuel. Une étude du contenu non-conceptuel comme celle présentée au chapitre 2.1 montre que le contenu non-conceptuel n'est pas un contenu de second ordre ou un contenu inférieur au contenu conceptuel, mais que sa fonction est différente de ce dernier.

En refusant de reconnaître le contenu non-conceptuel, Prinz poursuit en fondant le concept directement sur les données sensibles :

⁶ « *L'hypothèse de la priorité perceptuelle* : Rien n'est dans l'intellect qui n'est d'abord dans les sens (*nihil est in intellectu quod non fuerit in sensu*). »

⁷ « Le 'non-conceptuel' implique une distinction trop tranchée entre le perceptuel et le conceptuel. »

Concept Empiricism : All (human) concepts are copies or combinations of copies of perceptual representations ⁸. [Prinz, 2002, p. 77].

Le terme « copie » est à prendre en un sens métaphorique nous dit Prinz, au sens où une carte topographique est une copie d'un espace physique. Le seul problème est qu'une carte est une « *représentation* » et non pas une copie au sens où une version du penseur de Rodin peut avoir des copies de part le monde. Ces copies sont des « véhicules de la pensée » (*vehicles of thought*) [Prinz, 2002, p. 77]. Mais contrairement à un empirisme épistémologique, l'empirisme conceptuel de Prinz ne spécifie aucune forme de fiabilité et aucune condition de justification de l'assignation du concept. Et comme il ne détermine pas plus de conditions de signification ou d'usage, il se distingue de l'empirisme sémantique. L'empirisme conceptuel affirme simplement que les concepts sont construits comme des classes de représentations mentales d'origine perceptuelles, mais ne dit pas que ces représentations désignent effectivement quelque chose.

L'empirisme conceptuel délègue le contenu au sens, et présuppose qu'ils fonctionnent comme des détecteurs spécialisés (*dedicated input systems*), ce que Fodor appelle des *transducers*. C'est-à-dire qu'un capteur sensoriel de la couleur rouge activera le concept ROUGE sans que l'agent sache ni comment ni pourquoi. L'empirisme suppose également que le capteur en question est un bon capteur, c'est-à-dire qu'il est pertinent et fiable. Les concepts sont ensuite des copies de l'activation de ces capteurs, ils en dépendent donc entièrement.

The Modal Specificity Hypothesis : Concepts are couched in representational codes that are specific to our perceptual systems ⁹. [Prinz, 2002, p. 85].

L'inconvénient est que Prinz parle de « copie » de « représentation » de « code » sans jamais vraiment dire ce qu'il en ressort. Qu'est-ce qu'un « code représentationnel » sinon justement ce que l'on essaie d'expliquer lorsque l'on parle de concept ? Dire que les concepts sont des représentations

⁸ « L'empirisme conceptuel : Tous les concepts (humains) sont des copies ou des combinaisons de copies de représentations perceptuelles. »

⁹ « L'hypothèse de la spécificité modale : les concepts sont couchés dans des codes représentationnels qui sont spécifiques à nos systèmes perceptifs. »

mentales qui peuvent être activées par la mémoire de travail est une chose, l'expliquer en est une autre.

concepts are mental representations of categories that are or *can* be activated in working memory ¹⁰. [Prinz, 2002, p. 107].

Ce qui manque ici, est juste l'explication.

Prinz appelle ces représentations des « *proxytypes* », parce qu'ils valent comme des représentants, des délégués, des catégories qu'ils représentent.

I call representations “proxytypes”, because they stand in as proxies for the categories they represent ¹¹. [Prinz, 2002, p. 77].

L'individuation du proxytype par la catégorie, c'est-à-dire par la propriété, n'est pas suffisante, car elle ne permet pas de discriminer les concepts co-intensionnels ni les concepts co-référentiels. Pour palier ce défaut, Prinz suggère d'individuer le proxytype – le concept – par un ensemble de propriétés que par les propriétés détectées par le système sensoriel dédié, à la manière des « bouquets de propriétés » (*cluster of properties*) de Sellars [1992] :

I propose that proxytypes be individuated by sets of sets of properties rather than mere sets of properties. In particular, we can identify a proxytype by the set containing the sets of properties that are sufficient for causing the proxytype to exceed its critical detection threshold. For example, if a DOG proxytype includes the feature FURRY, BARKS, and FETCHES, and causing any two of these to be tokened is sufficient for tokening DOG, then that proxytype can be individuated by set containing four members : the set of appearances detected by FURRY and BARKS ; the set of appearances detected by FURRY and FETCHES ; the set of appearances detected by FETCHES and BARKS ; and the set of appearances detected by all of these representations.

This give us a way to compare proxytypes. Two proxytypes are identified if they detect the same appearance sets, and two proxy-

¹⁰ “les concepts sont des représentations mentales de catégories qui sont ou *peuvent être* activée dans la mémoire de travail.”

¹¹ “J'appelle ces représentations des “proxytypes”, parce qu'ils tiennent lieu de délégués pour les catégories qu'ils représentent.”

types are similar to the extent that the appearance sets they detect overlap ¹². [Prinz, 2002, pp. 275-276].

De fait les proxytypes sont individué par leur extension et non pas par leurs usages.

L'empirisme conceptuel tel qu'il est développé par Prinz est très séduisant et c'est l'une des théories du concept que l'on adopte – avec l'atomisme conceptuel – assez naturellement lorsque l'on se penche « naïvement » sur la question du concept. Un peu comme le réalisme interne est une métaphysique que l'on endosse lorsqu'on se prend les pieds dans le tapis et que l'on sait qu'on va se faire mal : « qu'importe que le tapis existe réellement ou non en dehors de la conception, ce que je sais c'est que je vais avoir mal ». L'empirisme conceptuel est une forme de bon sens.

Peut-être même que l'empirisme est un peu plus que ça. Il entend s'inspirer de la théorie des prototypes et de n'en garder que le meilleur. Le proxytype est un résumé des caractères de la catégorie, sans en passer par le calcul du degré de similarité, puisqu'il repose directement sur les capteurs sensoriels. Se faisant l'empirisme conceptuel explique l'ontogénèse et la possession du concept : un agent acquiert un concept grâce à l'expérience. Prinz accepte même une version faible de l'anti-innéisme et de l'anti-navisime en affirmant qu'il n'est pas impossible que l'agent possède « dès le départ » certains concepts primitifs. La dépendance du concept vis-à-vis de l'expérience permet d'expliquer le contenu cognitif du concept tout en rendant compte de l'effet de typicité.

Cependant la théorie des proxytypes développée par Prinz n'est pas très claire. Comment expliquer la publicité du concept ? Prinz nous dit qu'un

¹² «Je propose que les proxytypes soient individué par des ensembles d'ensembles de propriétés plutôt que par des simples ensembles de propriétés. En particulier, nous pouvons identifier un proxytype par l'ensemble contenant les ensembles de propriétés qui sont suffisantes pour déclencher le seul critique de détection du proxytype. Par exemple, si le proxytype CHIEN comprend les traits FOURRURE, ABOIEMENTS, et CHAIRS, et causer n'importe quel de deux de ces occurrences est suffisant pour occurer CHIEN, alors le proxytype peut être individué par l'ensemble contenant quatre membres : l'ensemble des apparences détectées par FOURRURE et ABOIEMENT ; l'ensemble des apparences détectées par FOURRURE et CHAIRS ; l'ensemble des apparences détectées par CHAIRS et ABOIEMENT ; et l'ensemble des apparences détectées par toutes ces représentations.

Cela nous donne un moyen de comparer les proxytypes. Deux proxytypes sont identifiés s'ils détectent les mêmes ensemble d'apparences, et deux proxytypes sont similaires en extension si les ensembles d'apparences qu'ils détectent se superposent.»

concept est public à partir du moment où le proxytype est partagé, mais comment savoir qu'un proxytype est partagé si le concept n'a pas de contenu objectif ?

D'autre part, le recours massif aux capteurs sensoriels pour expliquer le contenu du concept et sa fiabilité, ne rend compte que de l'expérience, et l'expérience, les illusions (cf. chapitre 2) l'ont montré, n'est pas uniquement conceptuelle. Ce que semble caractériser le proxytype est en fait le percept et non pas le concept, puisqu'il ne rend pas compte de l'usage du concept et des relations inférentielles entre les concepts.

3.6 Atomisme conceptuel

Atomisme conceptuel : *chaque concept est indépendant et autonome de tout autre aussi bien dans son usage que dans son application.*

L'*atomisme conceptuel* est la théorie des concepts mise en avant par Fodor [1998], et tout laisse à penser que la motivation première de cette théorie est de prendre le pendant inverse de toute forme de holisme. Le holisme est cette position qui affirme qu'un système constitue un tout qui ne peut pas être réductible à la somme de ses parties. Le holiste se défend en soutenant que les relations entre les parties influent sur le système lui-même et sur sa structure. Ces relations sont dépendantes du système et de ses composants, mais ne peuvent être tenues comme des éléments à proprement parler de ce système, mais le déterminent. L'image serait qu'au lieu de considérer un système sur la base d'une molécule composée d'atomes – métaphore utilisée par Frege pour rendre compte des énoncés complexes – le tout ressemblerait plus à une protéine composée à la fois d'acides aminés et de liaisons peptidiques. Rapportée dans le domaine du concept, l'interprétation holistique suppose que la possession de concepts par un agent cognitif détermine la possession qu'il a ou peu avoir d'autres concepts, ce qui implique une dépendance mutuelle des concepts. Développé de manière radicale, le holisme conduit tout droit au relativisme que l'on retrouve chez des auteurs tels que Kuhn [1962, 1970, 1996], Feyerabend [1980], Rorty [1997] ou encore Goodman [1984]. Et du relativisme, Fodor and Lepore ne veulent pas en entendre parler si ce n'est pour dire que c'est

la pire chose qui puisse arriver à un philosophe. Contre la théorie classique et contre la théorie-théorie des concepts, Fodor et al. [1980] soutient que les concepts ne peuvent se définir puisque cela signifierait que le concept défini dépendrait des concepts qui le définissent, et cela serait déjà admettre une forme de dépendance entre les concepts et donc une forme de holisme. Quant aux théories extensionnelles du concept, comme la théorie des exemplaires, la théorie des prototypes ou l'empirisme conceptuel, Fodor les rejette sous prétexte qu'il est absurde de constituer le concept à partir de rien, et que si on le construit à partir de l'expérience cela ne peut donner que des concepts dépendants de l'expérience, et donc relatif à l'agent, ce qui viole, selon Fodor deux conditions qu'il juge non négociable du concept, à savoir la publicité et la compositionnalité.

Pour lutter contre le holisme et l'empirisme – sous toutes leurs formes – Fodor [1998] développe une théorie du concept, l'*atomisme conceptuel* selon laquelle tous les concepts sont strictement indépendants et autonomes les uns des autres. Les concepts sont des *atomes* qui jouent le rôle de liaisons entre l'esprit et le monde. Et comme les concepts sont atomiques – c'est-à-dire autonomes et indépendants les uns des autres – chaque concept possède sa propre connexion avec le monde. La référence – l'objet du monde que le concept, dans l'esprit de l'agent, désigne – est déterminé de manière causale selon une loi :

Loi de référence causale du concept : le concept α réfère à φ du fait de deux propriétés de φ :

1. φ est φ
2. φ est la cause de toute occurrence actuelle ou possible de α pour dans un agent cognitif.

I continue to take for granted, as I've been doing all along of course, that semantic facts are somehow constituted by nomic relations. To a zero'th approximation, the fact that DOG means *dog* (and hence the fact that “dog” does) is constituted by a nomic connection between two properties of dogs : viz. *being*

*dogs and being causes of actual and possible DOG tokenings in us*¹³. [Fodor, 1998, p. 73].

La relation nomologique qui relie le concept à sa référence est constitutive du concept et est primitive, en conséquence Fodor ne l'explique pas plus avant. Le concept est donc comme un terme lexical, le terme « chien » dans le langage de la pensée, qui, tout comme le terme lexical 'chien' en français, désigne les chiens dans le monde. Le problème de la métaphore de la langue vernaculaire est que la signification d'un terme lexical est conventionnelle. Les locuteurs du français ou les experts de cette langue, ont décidé que le terme 'chien' désignerait les chiens. La relation entre les chiens du monde et le terme lexical n'est pas une définition, ni une dépendance entre différents concepts ou différents éléments du monde, mais simplement un lien rigide jeté entre l'esprit et le monde [Putnam, 1973, Kripke, 1972, 1980, Jacob, 2004]. Cette intentionnalité (avec un 't') du concept est primitive et provient du fait même qu'il y a quelque chose dans le monde qui est cognitif. La solution fodorienne est quasi cartésienne. C'est parce que nous pensons que nous pensons ainsi et nous pensons ainsi parce que nous pensons. L'intention ne s'explique pas, elle se constate :

having a concept (concept possession) is constituted, at least in part, by *being is* some sort of nomic, mind-world relation¹⁴. [Fodor, 1998, p. 121].

L'avantage de cette conception est qu'elle permet de dire que tout agent cognitif, en tant qu'il est un agent cognitif, possède déjà des concepts. Et comme les concepts ont leurs propres connexions au monde pour la détermination de leur référence, tous les agents possèdent des concepts préparamétrés, ce qui garantit la publicité du concept et permettent ainsi la communication entre les agents. Cela conduit à une forme d'innéisme ou de nativisme du concept, c'est-à-dire que l'agent possède déjà son lot de concepts en tant qu'il est un agent cognitif, et ces concepts il ne peut les modifier ni par l'influence d'autres concepts ni par l'expérience. Par

¹³“Je continue de prendre comme admis, comme je l'ai fait tout au long bien sûr, que les faits sémantiques sont d'une certaine manière constitués par des relations nomiques. Au degré zéro d'approximation, le fait que CHIEN signifie *chien* (et donc que le fait que 'chien' aussi) est constitué par une connexion nomique entre deux propriétés des chiens : à savoir *être des chiens* et *être les causes des occurrences actuelles et possibles de CHIEN en nous*.”

¹⁴“avoir un concept (la possession du concept) est constitué, au moins en partie, par *être* dans une espèce de relation nomique entre l'esprit et le monde.”

conséquent, la publicité et le partage des concepts ne sont pas le fruit de la communication mais c'est tout précisément l'inverse : la communication est le résultat de la publicité. Et par publicité il faut entendre ici le fait fortuit que tous les agents cognitifs, en tant qu'ils appartiennent à la même espèce, possèdent de manière naturelle, les mêmes concepts. L'agent évolue donc toujours, au niveau individuel, dans un solipsisme, mais la contingence fait que lorsqu'il parle à un autre agent, cela résonne dans la tête de son interlocuteur, comme ses propres concepts, et ainsi se fait la communication. Comme l'agent ne peut agir sur ses propres concepts, il ne peut se tromper dans leurs usages ou leurs applications. La conséquence négative de ce point est qu'il ne peut apprendre de nouveaux concepts, ou les enrichir, et quand un congénère lui communique quelque chose tout ce qu'il apprend il le sait déjà. L'atomisme de Fodor à des relents de la réminiscence platonicienne.

Fodor file la métaphore du concept comme terme lexical en allant jusqu'à dire qu'il y a un *langage de la pensée* (*language of thought*) [Fodor, 1974]. Il s'agit d'un langage avec sa syntaxe et sa sémantique, et en ce sens ce langage de la pensée se plie à la contrainte de générativité, à la Chomsky comme toute autre grammaire. Dans le langage de la pensée, le concept joue le rôle du terme lexical – du mot – mais à la différence d'un terme d'une langue vernaculaire étrangère qu'il faudrait apprendre, le langage de la pensée est comme une langue maternelle : on le parle sans savoir comment.

I'll going to claim, to put it roughly, that satisfying the metaphysically necessary conditions for having a concept *never* require the metaphysically necessary conditions for having any other concept ¹⁵. [Fodor, 1998, pp. 13-14].

En affirmant l'indépendance du concept, la relation causale directe entre le concept et sa référence et en rejetant toute notion de contenu non-conceptuel, Fodor est obligé d'admettre l'innéisme du concept.

Mais cette indépendance du concept va très loin, cela signifie qu'aucune relation privilégiée d'aucun type n'est nécessaire au concept. Cela signifie par exemple qu'il n'y a pas plus de lien entre les concepts BALEINE et MAM-

¹⁵“Pour le dire rapidement, je vais soutenir qu'en satisfaisant les conditions métaphysiques nécessaires pour avoir un concept ne requiert *jamais* les conditions métaphysiques nécessaires pour avoir aucun autre concept.”

MIFÈRE qu'il n'y en a entre les concepts BALEINE et POISSON ou BALEINE et JONAS. Un agent pourrait très bien posséder le concept BALEINE sans posséder le concept MAMMIFÈRE. Et si l'on essaie d'expliquer ce qu'est une baleine en disant que c'est un mammifère par exemple, cela n'avancera pas beaucoup notre interlocuteur car il ne pourra apprendre le concept BALEINE à partir du concept MAMMIFÈRE, les deux étant strictement indépendants. L'apprentissage étant banni, il ne reste plus qu'à l'atomisme conceptuel à dire que les concepts sont *primitifs*, c'est-à-dire qu'ils ne sont pas composés et qu'ils sont déjà livrés avec leurs propres références [Fodor, 1998, p. 121][Margolis and Laurence, 1999, p. 62].

Mais évidemment l'innéisme à ses limites : l'agent cognitif – du moins l'agent cognitif humain – n'est pas omniscient. Il est difficile de supposer qu'un agent sache tout et qu'il possède de manière innée un concept tel WIFI par exemple. Si un agent possède ce concept et que ce concept n'est pas inné, cela signifie qu'il est possible, d'une manière ou d'une autre d'acquérir un concept. L'atomisme conceptuel modère alors un peu ses ardeurs et affirme que certains concepts peuvent être composés à partir d'autres concepts. Il y a des *concepts primitifs* et des *concepts composés* formés à l'aide des premiers par composition, suivant le modèle de la générativité syntaxique de la linguistique. De fait, les concepts composés ne sont ni primitifs ni innés, mais se réduisent à des primitifs innés. D'une certaine manière cela rappelle la réminiscence platonicienne, où un concept est saisi à partir de concepts déjà possédés. Mais il ne faut pas confondre le concept et la connaissance. L'atomisme conceptuel n'est pas l'atomisme logique. L'atomisme logique soutient que les concepts atomiques sont nos connaissances du monde et sont en ce sens évaluables en termes de valeurs de vérité. Cette confusion peut se trouver chez Frege peut être, où le concept prend la forme d'un prédicat avec une valeur de vérité. L'atomisme conceptuel est psychologique. Les concepts sont des atomes qui ne sont que des représentations du monde, pour l'agent.

I am going to argue for a very strong version of psychological atomism : one according to which what concept you have is conceptually and metaphysically independent of what epistemic capacities you have. If this so, then patently concept couldn't be epistemic capacities ¹⁶. [Fodor, 1998, p. 6].

¹⁶“Je vais défendre une version très forte d'atomisme psychologique : une selon laquelle quelque soit le concept que vous avez il est conceptuellement et métaphysiquement

C'est-à-dire qu'un agent peut posséder le concept BALEINE sans nécessairement savoir que les baleines sont des mammifères. Penser que les baleines sont des mammifères est une connaissance qui implique l'usage de concept, mais qui est plus que le concept.

L'espace cognitif étant, selon l'atomisme conceptuel, structuré en concepts atomiques, les concepts complexes seront composés de ces concepts primitifs à l'aide d'un principe de composition basé sur une générativité. Il suffit de prendre deux atomes, de les coller ensemble pour faire un concept complexe. La compositionnalité s'explique presque d'elle-même. Pour faire le concept BOUTON DE MANCHETTE, il suffit de prendre le concept BOUTON et le concept MANCHETTE et de les accoler. La question est ensuite de savoir quel est le statut de ces concepts composés. La compositionnalité fodorienne ressemble à la combinatoire des idées proposée en son temps par [Leibniz](#), les termes composés devenant des abstraits.

D'un point de vue atomiste, les concepts sont des « *transducers* » des indicateurs de propriétés. Concept CHIEN se déclenche à chaque fois que se manifeste la propriété *canidité*, exactement comme la petite lumière sur mon ordinateur m'indique que la batterie est rechargée, ou la jauge d'essence indique la contenance du réservoir. En soit la jauge ou la lumière n'a pas de lien avec le niveau d'électricité ou d'essence, mais elles sont mises en relation telle qu'elles covarient de manière nomologique avec l'objet qu'elles représentent.

Cette covariation est implémentée par la loi de référence du concept telle qu'énoncée plus avant. Cette loi garantit que lorsque l'agent possède un certain concept, ce concept marche pour ce qu'il doit désigner. Et comme ce concept ne dépend que de l'architecture cognitive, cela signifie que tout agent d'une même espèce possède les mêmes concepts dessinés pour référer aux mêmes objets. Ainsi la publicité du concept est expliquée.

Pour accepter la notion de publicité selon l'atomisme conceptuel, il faut accepter l'innéisme et le nativisme qu'il suppose. Mais comme il est peu probable qu'un agent cognitif possède tous les concepts auxquels il recourt durant son existence de manière innée, comment expliquer l'acquisition de

indépendant des capacités épistémiques que vous avez. Si c'est le cas, alors évidemment le concept ne peut pas *être* une capacité épistémique.”

nouveaux concepts ? Acquérir un nouveau concept suppose l'enrichissement du vocabulaire conceptuel de l'agent, ce qui suppose qu'il puisse établir une loi nomologique de référence entre un objet du monde et un concept. Pour cela il y a deux solutions : ou bien admettre qu'il soit possible de composer des concepts complexes à partir de concepts simples et de leur assigner une référence propre qui ne dépend pas des références des composants ou bien de pouvoir assigner une référence à un concept suivant les occurrences perçues par l'agent. Mais ces deux options supposent ou bien que la référence du concept n'est pas nomologiquement prédéterminée, ou bien une forme de holisme et de relation entre les concepts. Ce que refuse explicitement l'atomisme conceptuel. Donc selon l'atomisme conceptuel, aucun nouveau concept ne peut intégrer le vocabulaire cognitif de l'agent, ce qui se pose en faux contre les données expérimentales de la psychologie du développement.

D'autre part, si les concepts sont réellement indépendants les uns des autres, comment est-il possible alors de justifier la hiérarchie des concepts ? Si l'une des fonctions du concept est de permettre de classer et d'organiser les éléments de l'environnement, cela signifie que les objets tombent sous les coups des concepts BALEINE, BOUTON DE MANCHETTE ou encore DINDE DE NOËL, ensuite que ces concepts tombent eux-mêmes sous les concepts ÊTRE VIVANT et ARTEFACT, et ainsi de suite. Mais la classification est structurée, elle indique une transitivité de propriétés dans un sens et d'extension dans l'autre, et donc une dépendance – au moins dans une dimension – des concepts. Si l'atomisme refuse cela, cela signifie qu'il refuse la catégorisation, et par là même nie l'une des fonctions du concept.

L'atomiste peut se défendre en soutenant que l'indépendance des concepts entre eux ne concerne que leurs conditions de possession et non pas leur référence. Mais quoi qu'il en soit, si le concept est une entité cognitive, c'est bien de la dépendance dans la possession dont nous parlons, et non pas d'un point de vue ontologique ou métaphysique. Ou bien alors l'atomiste pourrait soutenir que la hiérarchie des concepts et la transitivité intensionnelle qu'elle implique s'explique en termes de compositionnalité. Certes tous les concepts d'ordres supérieurs se réduisent alors aux concepts inférieurs, mais c'est bien le cas dans la classification – les mammifères sont la somme de tous les animaux qui allaitent leurs petits. De sorte que les concepts de haut niveau n'ont de référence que grâce aux concepts de bas niveau. Mais

alors cela signifie que le concept complexe n'a pas de référence propre, mais qu'elle dépend uniquement de ses composants. C'est-à-dire que le concept POISSON ROUGE voit sa référence empruntée à celle des concepts POISSON et ROUGE. Donc un coelacanthe badigeonné de purée de tomate est un poisson rouge, ce qui n'est évidemment pas le cas. À moins que dans la composition intervienne une restriction sur les propriétés mises en communs, mais cela implique qu'il est possible de mettre en commun les intensions des concepts, et le holisme ne dit pas autre chose. Si ce n'est pas le cas, si la mise en commun ne porte que sur l'extension, alors deux concepts vides seront confondus comme étant identiques, ce qui est absurde puisque les dahus ne sont pas des licornes. Que l'atomiste réponde que ce sont là deux concepts différents du fait de leur mode de désignation de l'objet, et que dans des mondes possibles dans lesquels il y a des dahus et des licornes on le distingue bien, on lui répondra alors que son attention se porte sur l'intension, et qu'alors il peut admettre des relations entre les concepts, du style « un dahu n'est pas une licorne », et en continuant ainsi, il empiète sur le holisme. L'atomisme ne s'en tire jamais seul.

3.7 Rôles conceptuels

Sémantique des rôles conceptuels : *posséder un concept c'est savoir l'utiliser et l'appliquer en fonction d'autres concepts.*

La sémantique des rôles conceptuels prend le contre-pied de l'atomisme conceptuel. Alors que celui-ci affirme la stricte indépendance et autonomie du concept vis-à-vis des autres concepts, la sémantique des rôles conceptuels défend la dépendance des concepts entre eux. La possession d'un concept entraîne, modifie ou influe sur la possession d'autres concepts. Il en va de même pour l'application et l'usage. Selon la sémantique des rôles conceptuels, posséder un concept c'est savoir l'utiliser, c'est-à-dire savoir l'appliquer aux objets auxquels il réfère dans le monde, et le mettre en relation avec d'autres concepts. L'usage correspondant alors aux inférences accessibles à partir de ce concept par l'agent cognitif. Un agent sera dit posséder le concept DINDE DE NOËL s'il est capable de montrer une dinde de Noël dans un rayon de supermarché, mais également s'il est capable de penser que les dindes de Noël ont un rapport avec le Père Noël et le 25 décembre.

Cela suppose que les concepts sont reliés entre eux. La sémantique des rôles conceptuels accepte donc une forme de holisme, ne serait-ce que partiel [Peacocke, 1992], c'est-à-dire que même si tous les concepts ne sont pas directement reliés inférentiellement les uns aux autres, certains groupes de concepts inter-reliés émergent qui correspondent à autant de domaines dont le pendant sont les théories naïves.

La sémantique des rôles conceptuels s'oppose à la référence causale du concept telle qu'elle est acceptée par l'atomisme conceptuel ou une téléologie de la référence à la Millikan, et ce parce que la sémantique des rôles conceptuels pense que l'application – la relation entre l'objet du monde et le concept – n'est pas suffisante pour rendre compte du concept. Selon la sémantique des rôles conceptuels, il faut comprendre la référence du concept à travers les relations entre les représentations de l'agent – sa ou ses conceptions – et son environnement. Si l'agent pense que les dindes de Noël sont des robots, alors le concept DINDE DE NOËL de l'agent ne sera pas relié au concept ÊTRE VIVANT mais à celui de FORCE DU MAL VENU DÉTRUIRE L'ESPÈCE HUMAINE, quand bien même dans le monde – la référence objective – les dindes seraient des êtres vivants. Cela n'est pas possible selon une théorie causale du concept, puisque la référence du concept ne dépend pas de la conception de l'agent, mais simplement du monde tel qu'il est.

Pour reprendre la métaphore de la protéine pour remplacer celle de la molécule, les acides aminés – joués ici par les concepts – sont reliés entre eux par des liaisons peptidiques jouées par des relations inférentielles. Ces relations sont extrêmement importantes puisque ce sont elles aussi qui déterminent les concepts. Un « même » concept peut être considéré comme différent suivant les relations qu'il permet.

L'importance mise par la sémantique des rôles conceptuels, sur les relations et l'interdépendance entre les concepts permet de rendre compte des concepts abstraits tels que les concepts mathématiques par exemple. L'optique est alors constructive [Hilbert and Bernays, 2001a] – ce qui n'empêche pas un réalisme – et chaque terme est déterminé non pas par une définition comme dans le cadre de la théorie classique mais par la position de ce concept dans la trame conceptuelle composée de tous les concepts possédés par l'agent. Posséder un concept devient alors la capacité d'utiliser ce concept

sur cette trame, un peu comme savoir jouer aux échecs c'est savoir déplacer les pièces du jeu selon une certaine stratégie. La sémantique des rôles conceptuels privilégie l'usage à l'application du concept. Et en un sens même, il est possible de *construire*, de *dérivée* ou de *postuler* la référence – l'extension – du concept, à partir de l'usage.

L'exemple de concept considéré par la sémantique des rôles conceptuels est celui de concepts dont il est difficile de rendre compte de l'extension par l'expérience. L'exemple canonique est celui de CONJONCTION. Selon la sémantique des rôles conceptuels, un agent sera dit posséder le concept CONJONCTION s'il maîtrise les règles d'introduction et d'élimination de la conjonction [Peacocke, 1992, p.6] :

$$\begin{array}{ll}
 (\wedge I) & A, B \vdash A \wedge B \\
 (\wedge E) & A \wedge B \vdash A \\
 (\wedge E) & A \wedge B \vdash B
 \end{array} \tag{3.8}$$

Un agent cognitif qui maîtrise le concept CONJONCTION ne peut manquer de tirer ces inférences, sinon cela signifie qu'il ne maîtrise pas, ou pas complètement ce concept. Ces inférences sont "inévitables" ("*primitively compelling*") selon Peacocke [1992, p. 6] ou des "régularités explicatives de base" ("*explanatory basic regularities*") selon Horwich [1994].

Deux interprétations peuvent être données à cela : soit l'agent cognitif, en tant qu'agent cognitif, ne peut manquer de tirer ces inférences du fait qu'il est un agent cognitif, autrement dit la structure même de la cognition donne ces inférences, tout comme l'architecture physiologique détermine les propriétés saillantes au niveau non conceptuel ; ou bien cela peut indiquer qu'il y a une forme de structure inférentielle *normative* pour la cognition. D'autres inférences pourraient être ou sont accessibles à partir du concept CONJONCTION mais la norme cognitive est que pour maîtriser le concept CONJONCTION l'agent doit être capable de tirer ces inférences-là. Quoi qu'il en soit, l'agent doit être en mesure de tirer des inférences et ces inférences, déterminant le concept, ne peuvent dépendre du concept. Cela suggère qu'il y a dans la cognition un « module logique » qui traite ces inférences. Probablement est-ce là le même module que celui qui intervient au niveau des proto-jugements [Rock, 1983, Crane, 1992], module qui reste à identifier, mais que les expériences psychologiques semblent identifier [Johnson-Laird and Byrne, 1991]. Cette conception est compatible avec une version évolutionniste du

concept, conception selon laquelle l'espèce humaine aurait acquis à travers le temps la faculté de représenter l'environnement et de tirer des inférences, et à partir de là, former des concepts. Si cela est le cas, cela signifie que d'autres espèces d'agents cognitifs pourraient avoir également développé cette faculté et donc également posséder des concepts. Concepts auxquels les humains n'auraient pas accès du fait de la spécificité du concept. Cela explique que si un martien venait à faire escale chez nous, il se pourrait très bien qu'il ait des concepts et une représentation du monde, mais que nous ne pourrions jamais communiquer avec lui dans sa conception [Davidson, 1984, 2001].

La sémantique des rôles conceptuels est avant tout une position philosophique et normative, il ne dit rien du module cognitif qui permet à l'agent de tirer des inférences, mais affirme simplement que si un agent possède tel ou tel concept alors il doit pouvoir tirer telle et telle inférence. Cette conception du concept explique le concept par l'usage, et l'application du concept dépend de son usage. La sémantique des rôles conceptuels s'inspire de la sémantique linguistique et de la pragmatique où la signification d'un terme dépend de son utilisation [Searle, 1958, Davidson, 1984, 2001, Marconi, 1997, Harman, 1987, Block, Peacocke, 1992].

La sémantique des rôles conceptuels rend compte de la manière dont l'agent cognitif utilise le concept et comment il peut acquérir d'autres concepts. La sémantique des rôles conceptuels explique par exemple que Paul pense que les baleines sont des poissons du fait de sa biologie naïve, et qu'il a modifié son concept de baleine en changeant la relation entre les concepts BALEINE et POISSON pour relier le concept BALEINE et MAMMIFÈRE. Les concepts étant reliés les uns aux autres, la nature holistique de la trame inférentielle ne fait aucun doute, même si ce holisme est partiel, délimité en domaines. Dire que les concepts sont reliés entre eux inférentiellement, c'est dire que les concepts sont dépendants les uns les autres, que la possession de l'un modifie et détermine la possession de l'autre. Mais cette dépendance conceptuelle doit être déterminée d'une manière où d'une autre, en fonction de la pertinence de la relation inférentielle. S'il est possible de maîtriser le concept MINISTÈRE DES FINANCES grâce au fait que $2 + 2 = 4$ ou de n'importe quelle tautologie, la relation entre ce fait et ce concept n'est pas des plus pertinentes. La pertinence de la relation inférentielle dépend du domaine ou de norme. Si le concept CONJONCTION est normé à partir des règles d'in-

troduction et d'élimination du connecteur, ce sont ces inférences que l'agent doit maîtriser et non pas d'autres. La sémantique des rôles conceptuels ne dit rien de la portée des inférences et de la trame inférentielle. Si l'on parle de norme, on se place au niveau de la publicité et de la communication du concept, mais il est tout à fait possible – même ne parlant de normes – de se limiter à un niveau purement idiolectal et à un usage privé du concept. La sémantique des rôles conceptuels explique que Paul pense que les baleines sont des poissons du fait de sa biologie naïve et de la relation qu'il a établi entre les baleines et les poissons, mais elle explique également qu'il devrait penser que les baleines sont des mammifères du fait que les experts en biologie marines affirment que les baleines sont des mammifères.

La faiblesse est peut-être de se focaliser sur l'intension du concept et de ne rien dire de son extension. La sémantique des rôles conceptuels laisse penser que l'extension du concept est déterminée par l'intension, sans en dire plus. En développant l'aspect de l'application du concept dans le cadre de la théorie des rôles conceptuels (cf. partie III), il est possible d'expliquer que la référence du concept peut être établie indépendamment de l'usage de celui-ci. Par exemple Paul peut penser que les baleines sont des poissons puis corriger son concept de BALEINE pour penser que ce sont en fait des mammifères, sans modifier l'application de son concept, c'est-à-dire en désignant toujours les mêmes éléments du monde avec celui-ci.

Si la possession du concept est déterminée par les inférences que l'agent peut tirer à partir de lui, cela suppose que l'agent doit être en mesure de tirer ces inférences. Une interprétation pourrait être que l'agent devrait simplement tirer ces inférences sans nécessairement savoir qu'il le fait. Mais à partir du moment où l'on reconnaît l'existence d'un niveau non-conceptuel il faut expliquer en quoi ces inférences sont différentes des proto-jugements. Cette différence tient au statut épistémique du concept, et donc représentationnel. Par conséquent l'agent qui maîtrise un concept doit pouvoir rendre compte des inférences qu'il fait pour utiliser ce concept, ce qu'Brandom [2000] appelle le « jeu des raisons ». Lorsqu'on demande à Paul ce qu'il entend par BALEINE, il doit pouvoir par exemple nous répondre « les baleines sont des poissons ». L'explicitation de l'usage du concept par l'agent explique la publicité du concept, puisqu'elle suppose la communication, le défaut et qu'elle semble impliquer un degré épistémique

fort, trop exigeant pour le concept. La communication du concept implique que le concept a deux niveaux : l'un au niveau de l'agent – un niveau privé – et un niveau public, et c'est cette version publique qui va être jugée par les autres agents. Pour expliquer ce niveau public, la sémantique des rôles conceptuels fait appel à un usage normatif du concept, sans expliquer quelles sont les relations entre l'usage privé et l'usage public du concept. L'usage privé, appelé également « contenu étroit » (« *narrow content* ») correspond à l'*usage cognitif* du concept, et l'usage public, appelé également « contenu large » (« *large content* ») à l'*usage canonique* du concept, tels qu'exposés au chapitre 4.

Le principal problème de la sémantique des rôles conceptuels tient à l'importance qu'elle accorde à l'intension du concept au détriment de l'extension. Si la sémantique des rôles conceptuels arrive à rendre compte des concepts abstraits, qu'en est-il des concepts empiriques? Le concept ne se définit pas uniquement par son usage, mais également par son application – sa référence. Il semble difficile de penser qu'un agent cognitif apprend le concept ROUGE uniquement en pensant que le rouge est une couleur et que les tomates sont rouges. L'expérience, le qualie semble important et même primordial pour l'acquisition des concepts empiriques, et de cela, la sémantique des rôles conceptuels, contrairement aux théories empiriques du concept, ne semble pas en rendre compte. En fait, du contenu référentiel du concept, la sémantique des rôles conceptuels ne dit rien, ou du moins pas directement. Peacocke [1992] tire le contenu du concept de ce qu'il appelle les « proto-propositions » et le « contenu scénario » (*scenario content*), qui correspond au découpage de l'expérience au niveau non-conceptuel. De manière un peu rapide, le contenu référentiel des concepts empiriques provient donc du contenu non-conceptuel. Cette conception ne peut tenir de manière aussi directe une fois admis l'autonomie du contenu non-conceptuel et le développement de sa structure, notamment en terme de *percepts* et de *proto-jugements*, puisque le contenu non-conceptuel direct est utilisé par le percept et non pas nécessairement par le concept.

La sémantique des rôles conceptuels ne se pose pas aussi clairement le problème de sa relation au contenu non-conceptuel, du fait que les notions de proto-jugement et de percepts ne sont apparues que récemment dans la littérature. Cependant, le problème rencontré par la sémantique des rôles

conceptuels relativement à la référence se comprend à partir des discussions au sujet de la divergence contextuelle de référence d'un terme linguistique. Le terme 'eau' n'a pas les mêmes conditions de référence et d'application sur Terre que sur Terre Jumelle; dans le premier cas 'eau' réfère à une substance désaltérante et incolore dont la structure chimique est H_2O , alors que dans le second cas, le terme 'eau' réfère à une substance désaltérante et incolore dont la structure chimique est XYZ [Putnam, 1975]. En imaginant un agent cognitif n'ayant aucune notion de chimie – c'est-à-dire à qui les compositions chimiques des deux substances ne seraient pas accessibles – il serait possible de penser que ses conditions d'usage des termes 'eau' sur Terre et sur Terre Jumelle soient exactement les mêmes bien que leurs applications divergent. Peut-être est-ce le cas des homonymes dans une langue, alors que les synonymes correspondraient à une divergence d'usage mais non d'application.

Les opposants farouches à la sémantique des rôles conceptuels lui reprochent son caractère explicitement holistique. Définir un concept par les relations inférentielles qu'il entretient avec d'autres concepts implique *de facto* qu'aucun concept ne peut être isolé, autonome et indépendant vis-à-vis d'aucun autre. La sémantique des rôles conceptuels se tient en symétrie inverse de l'atomisme conceptuel. Selon la sémantique des rôles conceptuels, un concept isolé n'aurait aucune utilité et ne serait jamais utilisé.

Cependant pour tout concept on peut montrer qu'aucune inférence n'est nécessaire ni suffisante pour le saisir. Prenez le concept ASPIRATEUR, on peut penser par exemple que le concept TUYAU ou le concept NETTOYER LA MOQUETTE peuvent être importants pour comprendre le concept ASPIRATEUR, et pourtant il est toujours possible de trouver une explication du concept ASPIRATEUR qui ne fasse intervenir aucun de ces concepts, donc ils ne sont ni nécessaires ni suffisants. Donc aucune inférence n'est vraiment indispensable pour posséder un concept, donc, disent les accusateurs, la sémantique des rôles conceptuels est fautive.

D'autre part, si comme le soutient la sémantique des rôles conceptuels un concept s'apprend en apprenant les relations qu'il entretient avec les autres concepts, alors cela implique que pour apprendre un concept il faut nécessairement déjà posséder un concept, mais comment l'agent possède-t-il le premier concept qui lui permet de posséder les autres concepts? Pour

cela il faut bien au moins qu'il y ait un concept qui soit isolé et qui permet d'apprendre les autres, et si cela est vrai alors la sémantique des rôles conceptuels est fausse.

Mais cela serait oublier d'une part le caractère normatif des inférences ou au moins de certaines inférences qui caractérisent certains concepts – par exemple les fondements des mathématiques de [Hilbert and Bernays](#) énoncent des caractéristiques relationnelles de certains concepts mathématiques que le concept doit avoir pour être tel qu'il est, est ces inférences sont normatives, ou bien rappelez vous la définition donnée pour la possession du concept CONJUNCTION. Et ensuite, c'est également faire mauvaise justice à la sémantique des rôles conceptuels que de dire qu'elle endosse nécessairement un holisme radical. [Peacocke](#), qui développe une théorie des concepts sur la base d'inférence, n'est partisan que d'un *holisme partiel* ou *local*, découpé en domaines d'application, et ne refuse pas que certains concepts soient déjà possédés naturellement et avant toute expérience, par l'agent. Les deux critiques contre le holisme de la sémantique des rôles conceptuels tombent donc.

Troisième partie

Bases pour une approche
dynamique du concept

Chapitre 4

Liminaires

L'hypothèse qui sert de point de départ de ce travail est qu'un *concept* est une entité mentale qui permet à un agent cognitif qui la possède de comprendre et de représenter son environnement. La notion de concept inclut donc aussi bien les représentations empiriques – les concepts recognitionnels par exemple – que théoriques – concepts formels par exemple.

Comprendre son environnement est à entendre ici au sens de représentation, c'est-à-dire, de correction : un agent cognitif comprend son environnement à partir du moment où il le représente, i.e. à partir du moment où il peut construire des croyances ou des connaissances à son sujet, former des jugements à son propos où élaborer des désirs et des stratégies pour agir sur lui. L'environnement est à prendre en un sens très large puisqu'il inclut aussi bien l'expérience perceptive que l'agent a du monde, mais aussi l'ensemble des représentations qu'il se fait de cet environnement même, ainsi que de lui-même, des autres agents et de leurs représentations et ainsi de suite si bien que la représentation de cet environnement contient en définitive tout ce sur quoi la pensée de l'agent cognitif peut porter.

Si les concepts sont des représentations alors ils réfèrent à quelque chose – par hypothèse autre chose qu'eux-mêmes. Par exemple, si CHIEN est un concept, alors il est l'entité mentale, dans l'espace cognitif de l'agent qui le possède – qui permet à l'agent qui le possède de référer aux objets du monde qui sont des chiens et uniquement à ces objets-là (moyennant peut-être une marge d'erreur pour traiter des cas tendancieux et des limites

épistémologiques de l'agent). Pour rendre compte des concepts il faut donc rendre compte du phénomène de l'intentionnalité : comment une entité mentale peut-elle désigner une entité du monde.

La question de l'intentionnalité des concepts soulève également la question de la sémantique des concepts : quelle valeur prend un concept lorsqu'il est appliqué à un objet du monde. Cette sémantique ne doit pas seulement rendre compte de l'adéquation ou non entre le référent et le référé (entre le terme et l'objet du monde) – e.g. x est un CHIEN si et seulement si x est un *chien* [Fodor, 1998] – mais également de l'*assentiment* que donne l'agent à cette application. En effet, le concept étant une entité mentale représentative, il est l'application par un agent de cette représentation à un objet du monde. Si l'objet du monde *Médor* est un chien, alors il est vrai que la propriété "être un chien" (C) est instanciée par *Médor*. Cependant cela n'est pas suffisant pour rendre compte du concept CHIEN. Le concept étant la représentation mentale dans l'espace cognitif de l'agent, il faut vérifier son application à l'objet pour l'agent en question qui possède ce concept. Paul peut croire pour une raison ou une autre que *Médor* est un chat. Le jugement que Paul formulera et qu'il tiendra pour vrai est que *Médor* est un chat. Or si *Médor* est effectivement – indépendamment de toute croyance par un agent – un chien, *Médor* ne peut pas être un chat. Il faut donc distinguer deux niveaux dans l'évaluation de l'application d'un terme à un objet. Le premier niveau est *propositionnel*, il concerne la valeur objective – indépendamment de l'agent – de cette assignation – et est vrai ou faux en fonction de l'état du monde, il s'agit donc d'une évaluation *métaphysique*. Le second niveau, par opposition, est *psychologique* ou *cognitif*, il concerne le *jugement* que l'agent fait, i.e. l'assentiment qu'il donne ou pas à l'assignation propositionnelle. Cette évaluation psychologique dépend de l'agent qui formule le jugement. Si Paul dit que *Médor* est un chat, il donne son assentiment à la proposition selon laquelle le terme 'chat' s'applique à *Médor*, c'est-à-dire qu'il tient pour vrai le fait que *Médor* est un chat. Si *Médor* est un chien, Paul se trompe. Cependant s'il croit qu'il est vrai que *Médor* est un chat et non un chien, ce jugement aura des répercussions sur d'autres jugements que Paul pourra faire, et même des jugements qui auront l'apparence de vérités métaphysiques mais qui seront pourtant psychologiquement faux. Par exemple si Paul croit que *Médor* est un chat et que de cette croyance il dérive la croyance que *Médor* est un mammifère, et s'il dit à Pierre que

Médor est un mammifère mais que Pierre croit que *Médor* est un chien, la proposition ‘*Médor* est un mammifère’ sera métaphysiquement vraie du fait que *Médor* est un chien et que les chiens sont des mammifères, en revanche, seul le jugement de Pierre sera correct, puisque celui de Paul se fonde sur une mauvaise raison, à savoir que *Médor* est un chat.

La sémantique des concepts doit donc rendre compte de la valeur des *jugements* et non pas uniquement celle des propositions assertées par les agents cognitifs.

D’autre part, la sémantique des concepts doit tenir compte de la possibilité laissée à un agent de suspendre son jugement ou l’assentiment qu’il donne à une proposition. Si les conditions de l’expérience ne sont pas suffisamment bonnes, Paul pour dire qu’il ne sait pas si *Médor* est un chien ou un chat. Si métaphysiquement cette indétermination de l’application d’une propriété à un objet n’a pas de sens – x est ou bien un chat ou bien un chien mais ne peut pas être les deux en même temps – par application du principe du tiers exclu – psychologiquement elle est tout à fait acceptable. Même si *Médor* est un chien, il se peut que Paul ne le sache pas parce qu’il n’est pas certain de discriminer correctement *Médor* ou bien parce qu’il ne sait pas très bien si, pour une raison ou une autre, *chien* peut se dire de *Médor* – e.g. pour une raison ou une autre Paul croit que *chien* s’applique plutôt à des animaux de couleur brune, et si *Médor* est noir, Paul pourra hésiter, même s’il n’est pas absolument convaincu que *Médor* n’est pas un chien, à dire que *Médor* est un chien.

Le fait qu’un agent suspende son jugement pour l’application d’un terme à un objet ne l’empêche pas pour autant de comprendre la proposition. Prenez l’exemple suivant :

Kiruna est au nord du cercle polaire. (4.1)

Imaginons qu’un agent ne sache rien de Kiruna, c’est-à-dire que s’il lui était demandé quelque chose à son sujet il ne saurait rien en dire. Cependant, en ignorant tout de Kiruna, cet agent – s’il comprend le français – pourrait comprendre la proposition (4.1). Il pourrait par exemple penser sur la base de (4.1) que Kiruna est le nom d’une localité, que si Kiruna est au nord du cercle polaire alors pendant au moins un jour dans l’année la nuit ne tombe pas, que si Kiruna est au nord du cercle polaire alors Kiruna est au nord

de l'équateur, etc. Et ces connaissances au sujet de Kiruna s'appliquent combien même l'agent ne pourrait rien dire simplement de Kiruna, elles viennent des liens entre les concepts qui apparaissent dans les jugements. Si l'agent relie le concept CERCLE POLAIRE au concept ÉQUATEUR par la relation asymétrique ' x est au nord de y ', tout ce à quoi s'applique x s'applique relation ' x est au nord de y '. La compréhension de la proposition (4.1) repose donc sur la compréhension d'une partie de celle-ci et de la reconstruction par supposition de la compréhension de l'autre partie de la proposition par déduction du sens de la proposition entière compte tenu des relations que l'agent peut établir avec des jugements qu'il possède déjà [Ranta, 1994, 6.5, p. 130]. Autrement dit l'analyse du jugement passe par l'analyse des autres jugements faits par l'agent. Même si l'agent est incapable de dire s'il est vrai ou non que Kiruna se situe au nord du cercle polaire, il peut tout de même comprendre que cette proposition a un sens et en dériver un certain nombre de choses – e.g. qu'il doit être vrai ou faux que Kiruna est au nord du cercle polaire – et ce, grâce aux concepts qu'il possède déjà. Les concepts que possède un agent influent donc sur l'utilisation des concepts par cet agent, soit en déterminant son application – comme dans le cas où l'agent ne veut pas dire que *Médor* est un chien parce qu'il croit que les chiens sont bruns et que *Médor* est noir – ou bien en permettant à l'agent de comprendre un énoncé dans lequel apparaît un terme qu'il ignore. Il faut noter dans ce cas que l'agent pourra dire qu'il ne sait pas si (4.1) est vraie ou non, mais il pourra tout de même tenir pour vrai la proposition 'Si Kiruna est au nord du cercle polaire alors Kiruna est au nord de l'équateur'.

La distinction entre le niveau métaphysique et le niveau cognitif permet de distinguer entre le concept et le terme lexical qui peut lui être associé. Il y a ainsi une différence entre le concept CHEVAL et le terme 'cheval'. L'une des différences les plus criantes est de dire que le concept CHEVAL peut être associé au terme 'horse' ou 'cavallo' alors que le terme 'cheval' et les termes 'horse' ou 'cavallo' appartiennent à des langues différentes et qu'ils ne sont pas identiques – par exemple 'cheval' est masculin alors que 'horse' est neutre. Mais la différence entre un terme et un concept concerne surtout leur traitement d'évaluation. Alors que le terme sera jugé au niveau *sémantique* en tant qu'il apparaît dans une proposition, le concept sera jugé

au niveau *pragmatique* en tant qu'il apparaît dans un jugement. De ce point de vue, contrairement à certains auteurs, le concept ne sera pas assimilé à un terme, mais peut-être à l'usage d'un terme :

Throughout much of this book, I have waffled on the question of whether I am talking about concepts or words. The two seem to be closely related. I can talk about children learning the concept of sheep, say, but I can also talk about their learning the word sheep. I would take as evidence for a child having this concept the fact that he or she uses the word correctly. On a purely intuitive basis, then, there appear to be considerable similarity between word meanings and concepts. And in fact, much of the literature uses these two terms interchangeably ¹.
[Murphy, 2002, 11, 385]

S'il est vrai que le concept est l'usage d'un terme, alors il est vrai qu'un agent utilisant un terme – e.g. un mot dans une langue, possède un concept qui est associé à ce terme – cependant il n'est pas nécessaire que tous les concepts aient nécessairement un équivalent lexical [Sperber and Wilson, 1986] – i.e. qu'à un concept corresponde un ou plusieurs termes – à moins évidemment de parler de manière métaphorique d'un langage de la pensée dans lequel à chaque concept correspondrait un et un seul terme.

La prise en compte de l'usage dans l'étude du concept, c'est-à-dire du contexte d'application, i.e. les *conditions d'assertabilité* d'un terme de propriété à un objet du monde, implique de prendre en considération les autres concepts possédés par l'agent, et l'attitude épistémique de celui-ci. Cela signifie que les concepts se situent dans une *trame conceptuelle* dans laquelle les concepts sont reliés les uns aux autres par des dérivations – ce qui conduit à un *holisme* –, et que d'autre part la sémantique utilisée ne sera plus bivalente comme celle de l'évaluation métaphysique de l'application au niveau propositionnel, mais au moins trivalente (oui, non, indéterminé) ou probabiliste.

¹ "Durant quasiment tout ce livre, j'ai hésité sur la question de savoir si je suis en train de parler de concepts ou de mots. Les deux semblent être intimement reliés. Je peux parler à propos de l'apprentissage du concept de mouton par les enfants, par exemple, mais je peux aussi parler à propos de leur apprentissage du mot mouton. Je voudrais prendre comme preuve qu'un enfant a ce concept le fait qu'il ou elle utilise le mot correctement. Sur des bases purement intuitives, alors, une similarité considérable apparaît entre les significations des mots et celles des concepts. Et en fait, la plus part de la littérature utilise ces deux termes de manière interchangeable!"

Que le concept soit compris au niveau du jugement et non au niveau de la proposition – en ce sens le concept n’est pas le composant des propositions mais des jugements ou des pensées – permet de le représenter sous forme prédicative : le concept apparaît alors comme l’assignation d’une caractéristique à un objet du monde pour comprendre celui-ci dans l’espace des représentations – suivant cette conception du concept, les concepts singuliers seront retranscrits sous forme de descriptions définies. Penser que *Médor* est un chien c’est déterminer l’objet ‘Médor’ en tant que *chien*, c’est en fait lui appliquer le caractère *chien* :

$$\textit{chien}(\textit{Médor}) \tag{4.2}$$

Cette forme est propositionnelle. (4.2) est vraie si et seulement si *Médor* est un chien. Elle devient une forme de jugement à partir du moment où un agent l’asserte (la forme propositionnelle est alors préfixée du signe d’assertion ‘ \vdash ’) ou la dénie (la forme propositionnelle est alors préfixée du signe de dénégation ‘ \neg ’) :

$$\vdash \textit{chien}(\textit{Médor}) \tag{4.3}$$

Si donc la forme prédicative n’est pas suffisante pour rendre compte du concept, puisqu’elle ne rend pas compte du jugement qu’en fait l’agent, elle permet cependant d’en simplifier sa compréhension. En effet un agent qui formulerait le jugement (4.3) serait dit posséder le concept CHIEN puisqu’il l’utiliserait pour décrire un objet, ici Médor. Que cette assignation soit correcte ou pas est une autre affaire, car quand bien même Médor ne serait pas un chien, l’agent en question tenterait bien de dire quelque chose – en se trompant donc manquant son but – à propos de cet objet qui est Médor, donc pour cet agent ‘*chien*’ représente bien quelque chose. On refrénera l’objection du laxisme d’une telle interprétation du concept selon laquelle tout ce qui représente et donc tout ce qui apparaît de manière consciente dans l’espace cognitif d’un agent tombe sous le coup de cette interprétation. Certes mais si nous avons défini le concept comme l’entité représentante dans l’espace cognitif d’un agent, c’est bien que tout ce que l’agent peut penser contient un concept. Qu’il y ait ensuite des différences au sein des concepts est une chose, mais donc l’objection porte, certes, sauf qu’elle n’en est pas une.

Si le concept CHIEN est l'assignation de la propriété *chien* à un objet du monde, alors cela signifie que pour n'importe quoi qui est un chien, l'agent réfère à cet objet par l'intermédiaire du concept CHIEN. L'assignation de la propriété est donc le point le plus important dans la compréhension du concept, plus que la prédication elle-même. Pour dire que Médor est un chien, il faut que l'agent possède le concept CHIEN, et ce concept est l'assignation de la propriété d'être un chien. Donc pour comprendre le concept on pourra extraire de (4.2) la propriété "être un chien" par abstraction :

$$\lambda\text{Médor}(\text{chien}(\text{Médor})) \quad (4.4)$$

L'abstraction procède de la sorte : si A est une formule et x est une variable d'individu, alors λxA est un terme d'abstraction.

Si $B[\varphi]$ est une formule et $T = \lambda xA$ un terme d'abstraction, $B[T]$ est la formule résultante de la substitution de tout $t \in \varphi$ par $A[t]$ dans $B[\varphi]$. Par exemple :

1	Paul possède Médor
2	Médor est un chien
3	<div style="border-top: 1px solid black; padding-top: 2px;">Paul possède un chien</div>

où 'Médor' est substitué par 'chien' du fait que MÉDOR \in CHIEN.

Lorsque $T = \lambda x(x \in Y)$ alors $B[T]$ est simplement $B[Y]$ alors Y est identique à l'abstraction $\lambda x(x \in Y)$. De fait pour toute substitution il suffit de considérer les termes d'abstraction [Girard, 1987, 3A, p. 176], [Takeuti, 1953]. Par exemple :

1	Paul possède un chien
2	chien est un chien
3	<div style="border-top: 1px solid black; padding-top: 2px;">Paul possède un chien</div>

'chien' étant ici l'ensemble des chiens du monde, incluant Médor, Rantaplan, etc. Ce chien n'étant pas plus avant spécifié, il n'est pas possible de dire que Paul possède Médor, mais simplement qu'il possède n'importe quoi qui est un chien.

Le concept agit comme une substitution d'occurrences d'une propriété dans l'assertion d'une forme prédicative, de fait le concept peut être compris comme un terme d'abstraction, i.e. un abstrait. On dira d'un agent cognitif qu'il possède le concept de CHAT s'il est capable d'appliquer la *félimité* à un objet x du monde, ou le concept ROUGE s'il est capable d'appliquer la *rougeur* à un objet x du monde, ou le concept KIRUNA s'il est capable d'appliquer *le fait d'être Kiruna pour quoi que ce soit qui est Kiruna* à un objet x du monde.

Le traitement du concept comme un abstrait permet de ne pas se préoccuper de ses occurrences lorsqu'on analyse son application. De fait si a et b sont des constantes d'individu liées dans A , alors le concept basé sur l'abstraction λxA , concept que l'on notera $\lceil \lambda xA \rceil$, s'appliquera aussi bien à a qu'à b . Si maintenant le concept $\lceil \lambda x\varphi \rceil$ s'applique à A , alors il s'appliquera à a et b dans A . Le caractère abstrait du concept permet donc la *contrainte de généralité* :

Thus, if a subject can be credited with the thought that a is F , then he must have the conceptual resources for entertaining the thought that a is G , for every property of being G of which he has a conception. This is the condition I call 'The Generality Constraint' ². [Evans, 1982, p. 104]

Generality Constraint : If a thinker can entertain the thought Fa and also possesses the singular mode of presentation b , which refers to something in the range of objects of which the concept F is true or false, then the thinker has the conceptual capacity for propositional attitudes containing the content Fb ³. [Peacocke, 1992, p. 31]

Si un agent utilise un concept il le fait dans une assertion qui représente quelque chose, c'est-à-dire dans l'assignation de l'application d'une propriété à un objet de l'environnement. Mais cela n'implique pas que l'agent sache de manière déterminée ce à quoi le concept réfère exactement. Un agent pourrait très bien endosser l'application selon laquelle les dahus sont des animaux présents dans le bestiaire folklorique alpin et donc que certaines

² "Ainsi, si un sujet peut être crédité de la pensée que a est F , alors il doit posséder les ressources conceptuelles pour recevoir la pensée que a est G , pour toute propriété d'être G de ce dont il a la conception. C'est cette condition que j'appelle 'la contrainte de généralité'."

³ "La *contrainte de généralité* : si un penseur peut recevoir la pensée Fa et qu'il possède également le mode de présentation singulier b qui réfère à quelque chose du rang des objets desquels le concept F est vrai ou faux, alors le penseur a la capacité conceptuelle pour les attitudes propositionnelles contenant le contenu Fb ."

histoires contées par de vieux savoyards parlent de dahus sans que cet agent sache exactement à quoi s'applique la propriété *dahu*. L'agent pourrait comprendre l'énoncé :

les dahus sont des animaux présents dans le bestiaire folklorique alpin (4.5)

car il comprendrait la proposition :

les x sont des animaux présents dans le bestiaire folklorique alpin (4.6)

et sous l'hypothèse charitable selon laquelle (4.5) est signifiante et que (4.5) est construite correctement – en respectant les contraintes génératives du français standard – à partir de (4.6), l'agent pourrait *construire* une référence au terme 'dahu' à partir des *conditions de vérité* de (4.6).

Dans les cas de synonymie – quand deux termes linguistiques différents réfèrent à un même objet – il se peut que le jugement de l'agent soit différent alors que la valeur métaphysique est la même :

L'étoile du matin est Vénus (4.7)

L'étoile du soir est Vénus (4.8)

même si métaphysiquement (4.7) et (4.8) sont équipotents, un agent pourrait penser que (4.7) est vrai mais que (4.8) est faux, ou même dire que (4.7) est vrai mais ne pas savoir si (4.8) est vrai ou faux, l'agent peut très bien dire qu'il ne sait pas et ne se prononce pas. Cette différence entre la valeur métaphysique et la valeur cognitive est due au monde d'évaluation des deux niveaux qui sont différents. Au niveau métaphysique la proposition est évaluée en fonction de son adéquation au monde ou non. Au niveau cognitif, l'évaluation se fait selon les raisons que l'agent a d'asserter ou non une proposition. Si pour une raison ou une autre l'agent croit que l'étoile du soir est Mars, il se trompera, certes, mais cela justifiera qu'il tienne pour faux l'énoncé :

L'étoile du matin est l'étoile du soir (4.9)

L'agent se trompe parce qu'il se fonde sur de mauvaises raisons pour soutenir et dériver son concept ÉTOILE DU SOIR. Mais cela ne concerne pas – du

moins pas directement – la valeur de vérité du terme ‘étoile du soir’, mais ses conditions d’assertabilité pour l’agent en question.

Le fait qu’un agent puisse *construire* et *postuler* la signification d’un terme qu’il ne connaît pas mais qui apparaît dans une proposition qu’il comprend s’il admet que celle-ci est signifiante – en lui appliquant un principe de charité – et en fonction d’autres occurrences de cette même forme que comprend d’agent. Par exemple si l’agent ne sait pas ce que signifie le terme ‘dahu’ mais postule que la proposition (4.5) a un sens parce que la proposition :

les chamois sont des animaux présents dans le bestiaire folklorique alpin (4.10)

a un sens, l’agent peut *postuler*, faire l’hypothèse, que le terme ‘dahu’ doit se trouver dans le domaine des objets qui permette de donner une valeur à la propriété “être un animal présent dans le bestiaire folklorique alpin”. Cette *présupposition* que fait l’agent pour comprendre la forme propositionnelle dont il ne comprend pas l’un des membres correspond à la postulation d’une référence – i.e. d’un objet auquel s’applique le terme en question et qui le rend vrai sous cette interprétation – mais cette référence est alors purement *construite*, c’est-à-dire que l’agent ne s’engage pas ontologiquement quand à l’existence de cet objet sur au niveau métaphysique, mais l’utilise au niveau cognitif sous forme d’une modalité hypothétique du type, “sous l’hypothèse qu’il y ait quelque chose comme un dahu alors selon l’énoncé en question, ce quelque chose apparaît dans le bestiaire folklorique alpin”, à partir de là, l’agent peut tenir pour vrai ou pour faux ou s’abstenir sur la vérité de la proposition (4.5) suivant les *raisons*, les indices et les données qu’il a pour la tenir pour vraie ou fausse, mais il s’agit alors d’une valeur *pragmatique* i.e. la valeur du jugement “je crois qu’il est vrai que les dahus sont des présents dans le bestiaire folklorique alpin”. L’agent *défère* à une autorité pour soutenir son jugement. Alors que dans le cas d’une évaluation métaphysique le critère est l’état du monde décrit.

Cela nous ramène donc à la distinction entre *terme* et *concept*. Un *terme* est un symbole dans un langage donné qui réfère à un objet ou à un ensemble d’objet dans le modèle d’évaluation de ce langage. Dans le cas des dahus, la référence métaphysique est l’ensemble vide dans le monde

actuel, puisque les dahus n'existent pas. Le *concept* est l'entité mentale qui renvoie à tout ce que l'agent relie le terme 'dahu'. Le concept correspond donc à l'usage que l'agent fait du terme. Comme il s'agit d'un usage, le concept apparaît dans un jugement, au sens où l'agent doit avoir des raisons pour utiliser ce concept plutôt qu'un autre. Au niveau cognitif donc il est possible de distinguer les licornes des dahus du fait que les mêmes choses ne sont pas nécessairement associées à ces deux termes, bien qu'au niveau métaphysique ces deux termes soient coextensifs et pointent l'ensemble vide. C'est parce que le concept se comprend au niveau cognitif qu'il est une entité *intensionnelle*, par opposition à l'évaluation métaphysique qui est *extensionnelle*.

La référence postulée pour le terme 'dahu' par l'agent pour évaluer l'énoncé (4.5), n'est pas une référence métaphysique, et il est difficile de soutenir qu'il s'agit même d'une référence cognitive. Tout ce que fait l'agent et de dire qu'en l'absence de preuves et de raisons il ne sait pas si ce terme réfère ou non a quelque chose dans le monde, mais si la proposition dans laquelle le terme apparaît et dont l'agent ne sait pas la valeur de vérité, est signifiante – i.e. est bel et bien une proposition avec une valeur déterminée – alors le terme en question doit avoir une référence. À partir de cette possibilité référentielle, si l'agent ne possède pas plus d'indices, il peut, à partir d'un principe de charité selon lequel un énoncé est signifiant et en l'absence de mention explicite du contraire, est vrai, l'agent peut *construire* un objet pour ce terme qu'il utilisera comme référé du terme. En somme tout terme considéré comme bien construit – et utilisé – est supposé avoir une référence. Mais en l'absence d'indices permettant de savoir quelle est cette référence, l'agent en postule une – il ajoute un objet possible au niveau métaphysique. Cette extension postulée correspond à l'extension possible des termes abstraits définie par Hilbert and Bernays [2001b, II, §1, p. 62]. Ces termes seront donc préfixés de l'opérateur ϵ de Hilbert. Cependant, nous l'avons vu avec le principe d'abstraction de Takeuti [1953], l'extension du terme peut être ignorée au niveau cognitif, puisque l'abstraction correspondante est suffisante pour utiliser le terme dans une inférence.

Par exemple, même si l'agent ne sait pas ce à quoi réfère le terme 'dahu' dans le monde – c'est-à-dire qu'il n'est pas capable de l'appliquer à un objet du monde – peut penser que ce terme réfère peut-être à quelque

chose et qu'un biologiste par exemple saurait mieux que lui trancher cette question, cependant en admettant l'usage du terme dans l'énoncé – sous cette hypothèse que l'énoncé dans lequel apparaît le terme a un sens – l'agent peut dériver des informations sur les dahus, par exemple que si ce sont des animaux alors ce ne sont pas des cailloux.

Le fait qu'au niveau du concept la référence métaphysique n'est pas requise et que la référence postulée peut suffire, est important, puisque cela permet de comprendre les situations dans lesquelles l'agent dit explicitement qu'il ne sait pas à quoi réfère un terme, dans une question par exemple. En demandant des précisions, l'agent demande à étayer son concept pour comprendre le terme, concept qui peut être très ténu et quasiment vide. Cependant il ne l'est jamais, au niveau cognitif il n'y a rien tel que des termes vides. Les concepts à partir du moment où ils sont utilisés – et donc à partir du moment où ils sont des concepts – possèdent des conditions d'assertion. L'agent en demandant des informations demande donc une explicitation, une clarification de l'usage du terme et éventuellement son application à un objet du monde, i.e. sa référence métaphysique. L'évaluation trivalente du concept (Vrai, Faux, Indéterminé) est différente de l'évaluation bivalente (Vrai, Faux) du terme. Cependant, comme l'agent assigne par présupposition une référence au terme qu'il ne connaît pas dans l'énoncé, il doit le faire sous la condition qu'il croit que cet énoncé est effectivement une proposition (c'est-à-dire que le terme réfère bien à quelque chose dans le monde mais qu'il ne connaît pas ce terme) : de cela découle le *principe de référence postulée* pour le concept :

$$(PRP) \quad \ulcorner \Gamma \vdash \alpha \urcorner_{a_i} \rightarrow \ulcorner \alpha : \epsilon \alpha \urcorner_{a_i}$$

Le principe de référence postulée (PRP) dit que pour tout concept α l'agent (noté ici a_i) postule que α a une référence (notée $\epsilon\alpha$), même s'il ne sait pas exactement laquelle. Si dans les cas où l'agent sait à quoi réfère le concept en question alors la référence postulée sera la référence réelle, i.e. métaphysique. Dans les cas où l'agent ne sait pas quelle est la référence réelle alors ce sera la référence postulée, i.e. construite.

Le problème posé par cette référence postulée – outre le fait qu’il soit possible de s’en passer en passant par les termes d’abstraction – est qu’il ouvre potentiellement la porte à une interprétation purement privée du concept. En effet, si l’agent postule une référence à un terme lorsqu’il est en situation d’ignorance, pourquoi devrait-il se trouver dans une situation de connaissance, alors qu’il lui suffit de postuler ce qu’il veut pour n’importe quel terme, et n’utiliser qu’un idiolecte ou un langage privé. À première vue rien n’empêche cela. Après tout un agent complètement isolé pourrait peut-être développer une représentation personnelle du monde qui ne collerait avec aucune autre, et il pourrait utiliser cette représentation pour décrire son environnement, forger des croyances et des connaissances dessus, y planifier des actions etc. Il ne communiquerait qu’avec lui-même sous forme solipsistique. Bien, mais tout l’intérêt du concept est justement qu’il rend possible la communication entre les agents, le transfert de connaissances, de descriptions de l’environnement, etc. C’est donc plus le cas où une communication et un apprentissage sont possibles, c’est-à-dire quand le concept est *public* dont il s’agit de rendre compte, plutôt que de parler de quelque chose dont on ne peut finalement rien dire.

Lorsqu’une communication marche, cela signifie qu’entre deux agents cognitifs, un concept est partagé – i.e. que le concept apparaît dans les espaces respectifs des deux agents. En fait il ne s’agit là que d’une supposition puisque si une communication n’échoue pas, cela ne signifie pas pour autant que les deux concepts (celui de chaque agent) soient identiques, mais simplement que rien dans l’usage d’un terme – le terme correspond alors au signe lexical dans le langage utilisé dans la communication associée au concept – n’est en contradiction avec l’usage qu’en font ces agents. Supposons par exemple que Pierre dise à François que Paul est célibataire. Pierre et François comprendront l’énoncé propositionnel “Paul est célibataire” à l’aide de leurs concepts respectifs de PAUL et CÉLIBATAIRE. Imaginons que dans cette discussion rien ne fasse échouer la communication. Il sera alors postulé que Pierre et François s’entendent sur le concept CÉLIBATAIRE, i.e. que ce concept apparaît dans leur espace cognitif respectif et que ces deux concepts sont identiques, similaires, sinon compatibles. Imaginons maintenant que Pierre dit à François que Paul est célibataire et parle du mariage de Paul. La communication échouera si pour François le concept

CÉLIBATAIRE est relié au fait que quelqu'un est célibataire s'il n'est pas marié. Dans ce cas François pourra demander des explications à Pierre en lui demandant ce qu'il entend par le terme 'célibataire' ou bien s'il parle du même 'Paul', bref, l'échec de la communication montrera que le concept CÉLIBATAIRE de Pierre n'est pas compatible avec le concept CÉLIBATAIRE de François. La communication, i.e. l'usage de termes dans un langage – notez ici qu'il est tout à fait possible que certains concepts ne soient reliés à aucun terme d'aucun langage, mais tout terme d'un langage suppose qu'il soit associé à un concept pour être signifiant, donc nous pouvons, par commodité, utiliser ces termes pour parler des concepts – suppose une publicité des concepts mais aussi une *compatibilité* de ceux-ci entre les différents agents.

Le caractère *public* du concept – pour son usage dans la communication entre agents – conduit à amender le PRP. Il faut ajouter le fait que l'agent est de bonne foi, c'est-à-dire qu'il entend postuler, pour un concept, la même référence que celle de son interlocuteur. Cette référence qui est celle qui devrait être la même que celle de l'interlocuteur est normative, elle correspond à la *référence canonique* du concept, c'est-à-dire la référence habituelle du terme associé au concept. Habituellement la référence du terme 'célibataire' regroupe l'ensemble des personnes non mariées et qui pourraient légalement l'être. Cet amendement est rendu par le *principe de référence postulée canonique* :

$$(PRPC) \quad B(\epsilon\alpha_{a_i} \doteq \epsilon\alpha_C)$$

L'agent croit que la référence qu'il postule pour le concept α est la même sinon compatible avec la référence canonique associée au concept α . C'est le réquisit pour comprendre la *publicité* du concept. Il se peut que ce principe soit utilisé de manière particulière, dans le cas de l'idiolecte par exemple, où l'agent décide de déterminer lui-même la référence canonique du terme. L'interprétation normale du PRPC devrait-elle selon laquelle l'agent croit qu'un expert – quelqu'un qui sait qu'elle est la référence métaphysique du terme associé au concept, pourrait le corriger. Ce principe est donc celui qui permet l'*apprentissage* et la *révision* du concept. Le cas de l'expert est précisément le cas de l'idiolecte. Un expert est celui qui dit que la référence qu'il postule pour un terme est ou devrait être la référence canonique de

ce terme. Prenez le cas d'un physicien qui parle de neutrinos. L'agent non physicien peut avoir une vague idée de ce qu'est un neutrino sans vraiment en être certain. La référence – et surtout l'usage du terme – qu'il va postuler, s'il en a besoin, se basera sur celle du physicien qui parle de neutrinos, de fait l'agent non spécialiste défère la référence au spécialiste, et lorsqu'il parlera de neutrinos il postulera que ce qu'il entend par là est compatible avec ce qu'en dirait un spécialiste. Mais le spécialiste lui défère sa référence à lui-même – si la modestie ne le ronge pas. Donc dans le cas où l'agent détermine lui-même la référence canonique [PRPC](#) reçoit une interprétation particulière, mais qui tient toujours, puisque l'agent entend bien référer à la même chose que l'usage canonique.

Dans l'exemple précédent de Pierre et François qui parle de Paul qui est célibataire, en cas d'échec de la communication, François demande à Pierre ce qu'il entend par 'célibataire' parce que la référence postulée par les deux est différente. Ce qui tranchera en définitive est ou bien la référence métaphysique – un célibataire c'est ça, et Pierre l'accepte – ou bien dans le cas où la référence métaphysique n'est pas connue – l'usage canonique du terme pour parler de certains objets du monde – généralement on entend ça par 'célibataire', au besoin un expert peut être consulté. Dans le cas de 'célibataire', si usuellement ce terme est utilisé pour parler des individus humains non mariés et qui pourrait l'être, alors cette interprétation servira de norme pour arbitrer la conversation, et Pierre devra, s'il est de bonne foi, réviser son concept de CÉLIBATAIRE. Rien n'empêche dans un contexte de communication qui n'a pas échoué que les deux concepts soient en fait incompatibles dans un autre contexte, mais alors, par charité, en l'absence de conflit, il sera supposé que l'agent postule une référence similaire ou compatible avec la référence canonique.

La référence postulée n'est que secondaire, elle sert à l'agent pour évaluer la proposition. Mais comme le concept apparaît au niveau du jugement – de la position assertive de l'agent vis-à-vis de la proposition – il est possible de comprendre le concept associé au terme en tant qu'abstraction – i.e. le terme doit avoir une extension, mais je peux me contenter de considérer la possibilité de son application – l'intension – et non son application elle-même – l'extension – pour asserter le terme. Et cela transparait dans l'ajustement des concepts dans le cas de l'échec de

la communication. François va demander à Pierre ce qu'il n'entend pas 'célibataire' parce que l'usage que fait Pierre de ce terme n'est pas le même que l'usage qu'en fait François. Cet usage concerne peut-être pour une part – pour les concepts recognitionnels peut-être – l'application de ces termes aux objets, mais surtout les relations entre les propositions dans lesquelles apparaisse ce terme. Lorsque François dit que les célibataires sont des individus non mariés, il rattache le concept CÉLIBATAIRE à la proposition "individu non marié et qui pourrait l'être", c'est-à-dire que là où apparaît le terme 'célibataire' peut se substituer la proposition "individu non marié et qui pourrait l'être". L'usage d'un terme se comprend alors comme les liens qui s'opèrent entre les concepts, et cela inclut les termes (simples ou complexes) associés au concept en question. Ces LIENS sont les *raisons* qu'à un sujet pour utiliser ce concept. Si François utilise le concept CÉLIBATAIRE pour parler de Paul c'est qu'il pense que Paul n'est pas marié. Les raisons agissent alors comme des *dérivations*, des *inférences* qui permettent et garantissent l'usage d'un concept dans un jugement, c'est-à-dire la valeur assignée à une proposition dans laquelle apparaît le terme correspondant au concept. Dans le cas où l'agent ne sait pas la valeur, i.e. quand il suspend son jugement, c'est qu'il pense qu'il n'a pas assez de raisons pour soutenir et garantir une valeur à son jugement. Lorsqu'il se trompe c'est qu'il se base sur des raisons qui ne permettent pas de dériver ces concepts ou bien qui ne sont pas des raisons compatibles avec les raisons canoniques qui justifient l'usage habituel du concept. Cette justification est donc elle-même normative dans la mesure où sa fonction est de régir la communication entre les agents. Il faut préciser également que cette communication n'est pas nécessairement lexicale, une communication non verbale, par exemple une action, un comportement, etc. pourrait tout à fait être *interprété* comme impliquant l'usage de représentations, et donc de concepts, et que suivant la compréhension par un tiers de ce comportement, celui-ci pourrait estimer que les concepts utilisés – si jamais des concepts sont utilisés pour réaliser cette action – par l'agent qui la réalise sont sinon identiques ou similaires, ou au moins compatibles avec les liens.

Si l'usage des concepts se base sur les raisons qui justifient la dérivation de ce concept compte tenu des jugements possédés par l'agent dans lesquels

ce concept apparaît, le concept se conforme au principe de dérivation :

$$(PD) \quad \ulcorner \alpha \urcorner_{a_i} \rightarrow \ulcorner \Gamma \vdash \alpha \urcorner_{a_i}$$

Si l'agent possède le concept α alors il a des raisons pour utiliser le concept α , i.e. il est capable de dire pourquoi il utilise le concept α tel qu'il le fait.

Si l'agent est de bonne foi alors il doit penser que les raisons qu'il invoque – raisons qu'il donnerait si quelqu'un lui demandait d'explicitier l'usage qu'il fait du concept – sont similaires, identiques ou tout du moins compatibles avec les *raisons canoniques* d'utilisation du concept – cela signifie qu'il pense qu'habituellement ce sont ces raisons-là qu'un agent doit avoir pour utiliser correctement ce concept – ce qui est rendu par le *principe de dérivation canonique*

$$(PDC) \quad B(\ulcorner \Gamma \vdash \alpha \urcorner_{a_i} \doteq \ulcorner \Gamma \vdash \alpha \urcorner_C)$$

L'agent pense que les dérivations qu'il utilise pour garantir l'usage qu'il fait d'un concept sont compatibles avec les dérivations canoniques pour ce concept. Cette clause est également normative et pragmatique, et son interprétation est également particulière dans le cas de l'idiolecte ou de l'expert dans lequel l'agent postule que ses dérivations pour garantir l'usage du concept sont et doivent être les dérivations canoniques pour déterminer l'usage canonique de ce concept. Les raisons doivent être explicitables, il est possible de questionner un agent sur l'usage de ses concepts et celui-ci doit se justifier ou bien se confronter à l'usage canonique des concepts. Dans les cas de disputes entre experts où chacun veut défendre sa référence ou dérivation comme étant canonique, le débat est tranché – si jamais – par la qualité, l'efficacité et la solidité des meilleures raisons apportées pour dériver un concept. Mais cela ne marche que si l'agent est de bonne foi – on qualifiera un agent de bonne foi capable de donner des raisons d'agent rationnel. Un agent de bonne foi doit donc croire qu'il est possible – s'il n'est pas expert – que la référence qu'il postule pour un concept puisse être différente de la référence canonique postulée pour ce concept. C'est-à-dire que l'agent accepte la possibilité qu'il puisse se tromper dans l'application d'un concept, cela est transcrit par le *principe d'incompatibilité possible de référence* :

$$(PIPR) \quad B\Diamond(\ulcorner \epsilon \alpha \urcorner_{a_i} \not\equiv \ulcorner \epsilon \alpha \urcorner_C)$$

de même que l'agent doit accepter que les raisons qu'il donne ne sont pas nécessairement compatibles avec les raisons canoniques pour les dérivations : *principe d'incompatibilité possible de dérivation* :

$$(PIPD) \quad B \diamond (\ulcorner \Gamma \vdash \alpha \urcorner_{a_i} \not\equiv \ulcorner \Gamma \vdash \alpha \urcorner_C)$$

Comme la référence peut être ignorée pour l'usage du concept si celui-ci est compris en tant qu'abstraction, le principe le plus important est donc le *principe d'incompatibilité possible de dérivation* (PIPD). Ce principe affirme qu'un agent cognitif reconnu comme rationnel, i.e. comme utilisant de bonne foi des concepts qu'il pourrait partager avec un autre agent, accepte d'expliquer son usage du concept et dans le cas où il n'est pas expert qu'un expert pourrait éventuellement le corriger si son usage était incompatible avec l'usage canonique utilisé dans la communication. En fait ce principe s'applique même dans le cas de l'expert lorsque son usage du concept est refusé, soit par un autre expert, soit par un profane. L'expert doit être en mesure de dire pourquoi son usage du concept est, et donc est conforme, à l'usage canonique. Il faut également noter que les raisons invoquées pour utiliser le concept ne sont pas nécessairement bonnes, et pourrait tout à fait être taxées d'irrationnelles. Cependant elles servent tout de même de justification pour dériver et utiliser le concept en question. Et il est également possible d'invoquer des raisons jugées non formulables, par exemple l'évidence des faits. Mais alors sur le plan formel, cela apparaîtra tout de même comme une dérivation, et donc une justification donnée.

Un agent cognitif sera dit rationnel au regard de son utilisation des concepts s'il accepte que les raisons qu'il invoque pour utiliser ses concepts ne soient pas nécessairement identiques, similaires ou compatibles avec les raisons canoniques d'utilisation de ces concepts, ce que rend le *principe d'agent cognitif rationnel*

$$(PACR) \quad B \diamond (\ulcorner \epsilon \alpha \urcorner_{a_i} \not\equiv \ulcorner \epsilon \alpha \urcorner_C)$$

Le **PARC** n'est pas une définition suffisante de la rationalité, mais il doit être supposé pour rendre compte des concepts. En effet, si ce qui précède est vrai, alors le concept est une entité mentale représentative qui permet l'usage de termes dans des propositions à travers des jugements.

À partir de là il faut expliquer maintenant comment fonctionnent les dérivations, rendre compte de la publicité du concept – comment une entité dans l'espace cognitif d'un agent peut être corrélée à une entité dans l'espace cognitif d'un autre agent –, de l'apprentissage des concepts et de leur révision.

Chapitre 5

Inférence

Si les concepts représentent quelque chose et s'ils sont les constituants de la pensée alors les concepts doivent être reliés les uns aux autres d'une manière ou d'une autre. En effet si par exemple le concept ANIMAL permet de représenter quelque chose, mettons l'objet o_1 de l'expérience et que le concept CHAT permet également de représenter l'objet o_1 en question, alors les concepts ANIMAL et CHAT ont un lien l'un à l'autre. Remarquez que cela vaut également pour les concept ANIMAL et AUTOMOBILE relative à l'objet o_1 en question, puisque s'il est vrai par exemple que l'objet o_1 est un animal alors il est faux que l'objet o_1 est une automobile, ce qui nous permet de mettre en relation les concepts ANIMAL et AUTOMOBILE au moins en ce qui concerne l'objet o_1 .

Lorsqu'on analyse le rapport qu'il peut y avoir entre deux concepts concernant un objet donné par exemple, on analyse en fait la relation qu'il y a entre deux – ou plusieurs – formes de fixation d'une pensée. Dire que l'objet o_1 vérifie le concept ANIMAL c'est dire que la pensée "cet objet est un animal" est fixée. C'est-à-dire que l'agent cognitif énonçant cette pensée tient le fait que l'objet o_1 en question est un animal. Il y a plusieurs manières possibles de fixer une pensée, l'agent peut parfaitement avoir la pensée que l'objet o_1 est un animal sur la base par exemple d'une expérience de o_1 et alors la pensée sera fixée à partir du contenu de cette expérience et pour une part au moins – sinon toute – cette expérience est non-conceptuelle . Mais l'agent cognitif peut également posséder la pensée que l'objet o_1 est un animal sur la base du fait que l'objet o_1 est un chat par exemple.

Le processus qui permet de fixer une pensée – i.e. de la tenir pour vraie – à partir d'autres est appelé *inférence*. Dans l'exemple, l'agent cognitif infère la pensée que l'objet o_1 est un animal en vertu du fait de la pensée que l'objet o_1 est un chat. Formellement l'inférence est du type :

1	Les chats sont des animaux
2	L'objet o_1 est un chat
3	L'objet o_1 est un animal

Cela s'explique par le fait que *factuellement* les chats sont compris parmi les animaux, autrement dit l'ensemble des chats est un sous-ensemble des animaux. On reconnaîtra la maîtrise du concept CHAT à un agent cognitif si celui-ci par exemple est capable de tirer la conclusion dans l'inférence précédente. Pour se faire il faut que l'agent cognitif possède la capacité de tirer des conclusions à partir de prémisses, ce que [Marconi \[1997, p. 2\]](#) appelle-la *capacité inférentielle* (*inferential ability*). Posséder la capacité inférentielle c'est avoir accès à un réseau de connexions entre les concepts.

Mais cela implique également que les concepts puissent être reliés les uns aux autres dans un réseau inférentiel. La structure même de ce réseau impose une contrainte importante sur les concepts eux-mêmes, en particulier cela signifie que les concepts doivent se conformer aux structures formelles de l'inférence – cela ne signifie pas pour autant bien évidemment que les concepts respectent nécessairement les structures formelles de l'inférence telle que nous les énonçons, mais cela signifie que ces structures formelles sont un outil permettant de représenter ces contraintes sur les concepts, nous ne sommes pas en train de soutenir que la cognition humaine est régie effectivement par les règles de la logique formelle, mais que celle-ci permet, dans une mesure au moins, d'en rendre compte.

Les concepts doivent donc avoir des propriétés logiques [[Sperber and Wilson, 1986, 2, p. 68](#)], ils doivent par exemple être en mesure de s'impliquer les uns les autres, e.g. le concept CHAT implique le concept ANIMAL i.e. si o_1 est un chat alors il est aussi un animal, ou au contraire d'entrer en contradiction les uns les autres, e.g. le concept ANIMAL est contradictoire avec le concept AUTOMOBILE i.e. si o_1 est un animal alors il n'est pas en même temps une automobile. Il est évident également que de dire que les concepts ont des propriétés logiques ne signifie pas que les concepts n'ont

que des propriétés logiques, le fait par exemple d'être une entité mentale n'est pas une propriété logique du concept.

Nous avons dit qu'il était possible de fixer une pensée à partir du contenu non-conceptuel de l'expérience ou bien à partir du contenu représentationnel conceptuel. La différence entre ces deux types de fixation de la croyance se retrouve dans le type d'inférence. La fixation sur la base du contenu non-conceptuel se fait sous la forme de systèmes dédiés de détection (*input system*) ce qu'on appellerait aussi des systèmes de véhicule d'information ou des détecteurs [Dretske, 2003, p. 38][Dretske, 2000, p. 67]. La fixation de la pensée à partir du concept, contrairement à la fixation de la pensée à partir de l'expérience non-conceptuel, et qu'alors que la seconde est dépendante du contenu présent de l'expérience et du contexte de celle-ci, la première peut se baser sur le contenu des concepts indépendamment de l'exemplification de celui-ci au moment de l'inférence. Autrement dit à partir de l'expérience on peut voir un chat (sans nécessairement voir *que* c'est un chat) si et seulement si il y a un chat dans notre champ visuel dans des conditions appropriées de vision mais pas autrement. Alors qu'il est possible de dire qu'un chat est un animal quand bien même nous n'aurions pas de chat dans le domaine présent de notre expérience. L'induction part des instances et est donc relative à celles-ci ce qui ne confère qu'un statut probable à sa conclusion alors que la déduction part de prémisses et en dérive par des règles une conclusion, cette dernière ayant une valeur déterminée relativement aux prémisses et à la règle. L'importance alors est celle de la règle de déduction. L'avantage des règles de déduction est de séparer la forme du concept de son contenu. Les formes utilisables d'inférences seront connues sous forme de schémas valides indépendamment du contenu des termes qui apparaissent dans l'inférence. Chaque inférence à partir de concept dépend évidemment du contenu des concepts en question mais en s'inscrivant dans un schéma abstrait de déduction. La dissociation entre la forme et le contenu de l'inférence permet de stocker plus facilement en mémoire les inférences déjà faites en associant des concepts en fonction de règles utilisables par ailleurs et d'élargir les inférences possibles à l'ensemble possible des termes qui peuvent apparaître dans un tel schéma. De fait et contrairement à ce que permet l'expérience, tout nouveau concept ou terme peut s'inscrire directement dans une inférence sans devoir auparavant

restructurer l'ensemble des concepts et des termes stockés en mémoire par l'agent cognitif. D'autre part les schémas d'inférences ne dépendant pas du contenu mais de la forme, i.e. étant abstraits, permettent de certifier de l'exactitude de la pensée dérivée à partir de l'exactitude des prémisses employées et par la règle de déduction. De fait ils permettent par conséquent de déterminer les inconsistances et les contradictions dans les inférences non valides pour toute représentation conceptuelle possible. Par exemple, en partant du donné de l'expérience rien ne permet de dire que si o_1 est détecté par le système dédié à la saisie des animaux mais pas par celui dédié à la saisie des automobiles que ce qui est détecté par le système dédié à la saisie des animaux n'est jamais détecté par le système dédié à la saisie des automobiles. Par contre par la déduction si nous disposons de prémisses du type :

1. Les animaux sont des êtres vivants.
 2. Les automobiles sont des artefacts.
 3. Les artefacts ne sont pas des êtres vivants.
-
4. \therefore Une chose ne peut être à la fois un animal et une automobile.

L'introduction d'une contradiction suppose que les concepts soient confrontés entre eux selon ce qu'ils représentent, tant du point de vue extensionnel que du point de vue intensionnel. Et pour se faire il faut les comparer sur l'ensemble par exemple de leurs occurrences et non pas seulement celles disponibles dans l'expérience présente. Comparez avec une expérience qui activerait le détecteur dédié à la saisie des animaux mais pas celui dédié à la saisie d'une taille de plus d'un mètre, si une déduction ou une généralisation était faite sur cette base elle conclurait qu'il n'est pas possible d'une chose soit à la fois un animal et fasse plus l'un mètre par exemple. Or ces deux concepts, ANIMAL et MESURER PLUS D'UN MÈTRE ne sont pas contradictoires sans pour autant qu'ils soient exactement co-extensionnels ou co-intensionnels. La déduction permet de dire que certains et certains seulement des animaux mesurent plus d'un mètre mais que d'autre mesurent moins.

Cette comparaison et la relation qu'elle fait ressortir entre les concepts amène à dire que les concepts doivent se conformer aux schèmes inférentiels mais également qu'il soit possible de discriminer deux concepts quant à

ce qu'ils représentent c'est-à-dire aussi bien à leur extension, qu'elle soit objective (le contenu objectif, ce que le concept représente d'un point de vue général) ou subjective (le contenu cognitif, ce que l'agent cognitif associe au concept en fonction de son expérience et des autres concepts qu'il possède), que de leur intension (la propriété associée au concept), cela suppose qu'il soit possible donc d'individuer et d'identifier les concepts et cela suppose que les concepts soient des entités abstraites.

Les prémisses et la conclusion d'un schème inférentiel sont composées de concepts. La prémisse (1) par exemple est composée au moins des concepts ANIMAL et ÊTRE VIVANT puisque ces concepts peuvent apparaître séparément dans des inférences et que toutes les inférences permises par l'un ne sont pas nécessairement accessibles à l'autre. De manière évidente toutes les inférences possibles à partir du concept ÊTRE VIVANT ne sont pas toutes dérivables à partir du concept ANIMAL. Il est même possible de considérer que "les" dans la prémisse renvoie au concept LES puisqu'il apparaît aussi dans la prémisse (2) sans qu'aucun des deux autres concepts que l'on vient de citer n'y figure. La question de savoir s'il faut rattacher le terme "les" au concept LES ou bien au concept PURIEL ou MARQUE DU PLURIEL par exemple est une autre histoire qui relève de la dimension *lexicale* du concept [Sperber and Wilson, 1986, 2, p. 89][Murphy, 2002, 11, p. 385], c'est-à-dire aux termes de la langue (quelle soit vernaculaire ou formelle) associés au concept. D'autre part les concepts utilisés dans (1) et (2) sont les mêmes qui apparaissent dans (4).

5.1 La distinction des concepts

Pour pouvoir entrer dans une trame ou un réseau inférentiel un concept doit pouvoir être distingué, c'est-à-dire individué et identifié par rapport aux autres concepts. L'identification du concept se base sur sa généralité : c'est dans la mesure où l'on peut délimiter le concept tant en extension qu'en intension qu'il est possible de le séparer des autres. La généralité du concept permet de savoir à quoi s'applique le concept, i.e. de déterminer son extension, l'ensemble des objets qui satisfont le concept, et la propriété associée au concept que les objets qu'il subsume doivent avoir pour tomber sous le concept, i.e. son intension. Le concept s'appliquant à l'ensemble des objets vérifiant la propriété à laquelle il est associé dans les mêmes conditions

d'application, un agent cognitif possédant un concept F dans la pensée Fa , c'est-à-dire lorsque le sujet a la croyance que l'objet a a la propriété F , par exemple que o_1 vérifie le concept CHAT et que cet agent cognitif a accès également à l'objet b à travers une expérience, c'est-à-dire un *mode de présentation* [Peacocke, 1992, 2.2, p. 42], [Recanati, 1993, §6] et que b réfère à un objet dans la portée du concept F , c'est-à-dire dans le domaine dans lequel F peut être évalué comme étant vrai ou faux, alors l'agent cognitif a la *capacité conceptuelle* de fixer des croyances ayant pour contenu Fb . L'agent cognitif possédant un concept et ayant accès à des occurrences possibles de ce concept doit être capable de forger des pensées en appliquant le concept aux occurrences qui le vérifient. Cela permet de définir la généralité du concept eut égard à sa portée.

La contrainte de généralité stipule donc que le concept est identifié par son extension, c'est-à-dire que l'agent cognitif doit pouvoir appliquer le concept aux différentes instances de celui-ci qui apparaissent dans les mêmes conditions d'application et du point de vue de leur intension : toute instance présentant la propriété associée au concept doit permettre l'application du concept à ces instances.

L'énoncé sous cette forme de la contrainte de généralité conduit à penser le concept en terme de prédicat. Mais nous verrons qu'il faudra clarifier cette conception afin de distinguer clairement l'intension, l'extension et l'application conceptuelle du concept.

La contrainte de généralité i.e. le caractère abstrait du concept lui permet de représenter, c'est-à-dire de garder la référence à l'objet qu'il caractérise lorsque l'objet n'est plus accessible à l'expérience. Cela suppose qu'il est possible de distinguer les instances du concept pour ce qui est de son extension et la ou les propriétés qui lui sont associées pour son intension de manière autonome c'est-à-dire sans avoir à une comparaison dans l'expérience entre plusieurs objets donnés simultanément. C'est la différence qu'il y a entre la *présentation* par l'expérience dans le mode de présentation et la *représentation* au niveau de la généralité du concept.

Au niveau non-conceptuel une forme de discrimination des propriétés est possible à partir d'une individuation mais qui n'est pas suffisante pour identifier la propriété. La représentation, le fait de pouvoir individuer une *même* propriété dans différents modes de présentation suppose de distinguer

les propriétés. L'identité nécessaire ici s'avère requérir un principe coûteux à mettre en place, à savoir le principe d'identité.

5.2 L'inférence comme maîtrise du concept

Si le concept apparaît dans la représentation du monde et si les concepts sont reliés entre eux par des liens inférentiels, alors la représentation du monde, c'est-à-dire le système de croyances et de connaissances d'un agent cognitif apparaîtra dans la maîtrise des liens inférentiels opérants entre les concepts qui structurent cette représentation. Si un agent cognitif est capable de tirer certaines inférences à partir de concepts, i.e. de passer d'un concept à un autre, il sera reconnu que cet agent possède et maîtrise ce concept. Par exemple si un agent cognitif possédant la croyance :

$$\text{Paul est célibataire} \quad (5.1)$$

et que cet agent est capable d'inférer à partir de (5.1) la croyance :

$$\text{Paul n'est pas marié} \quad (5.2)$$

il sera reconnu que cet agent maîtrise le concept CÉLIBATAIRE puisqu'il est capable de réaliser une inférence valide à partir de ce concept. En effet la mise en relation de la croyance (5.2) avec la croyance (5.1) suppose que l'agent cognitif soit passé par un raisonnement du type :

1. Les célibataires ne sont pas mariés.
2. Paul est célibataire.

3. \therefore Paul n'est pas marié.

La prémisse (1) utilisée correspond à la définition de ce qu'est qu'être célibataire. Une définition indique simplement qu'un terme introduit dans un système donné à la même dénotation d'un terme déjà connu dans le système et que l'un peu être utilisé pour d'autre pour toute substitution extensionnelle en contexte comparable. Autrement dit aucun célibataire n'est marié

et toute personne non mariée est célibataire. La propriété qui caractérise le fait d'être célibataire recoupe celle d'être non marié. L'indication du système relatif dans lequel s'opère la définition est importante. Il serait possible par exemple de mettre en relation le terme "célibataire" et le terme "bachelor" en disant que ces deux termes ont la même extension dans la mesure où ils caractérisent tous les deux Paul, mais comme ces deux termes apparaissent dans des structures langagières et lexicales différentes, l'un dans le français l'autre dans l'anglais ces deux termes ne seront équivalents que relativement à un système tiers comprenant ces deux systèmes par exemple dans un exercice de traduction où l'énoncé de la croyance "Paul is a bachelor" sera mis en correspondance avec la croyance "Paul est célibataire". Il est possible que si la croyance "Paul is a bachelor" était rendue directement dans une traduction par exemple par "Paul n'est pas marié", cette traduction serait jugée moins rigoureuse que celle la rendant par "Paul est célibataire". Cela soulève la remarque qu'il s'agit de distinguer les termes linguistiques associés à un concept et ce concept lui-même. Il est tout à fait envisageable par exemple que les termes "célibataire", "non marié", "bachelor" et "unmarried" (bien que bachelor soit restreint au sexe masculin en anglais) se rattachent tous au même concept mettons CÉLIBATAIRE. Mais le terme linguistique est une manière commode d'avoir accès au concept [Murphy, 2002, 11, p. 385]. Sans doute n'est-ce pas la seule manière – les analyses des conceptions naïve chez les enfants pré-verbaux s'attachent bien à l'étude des concepts [Bloom, 2004].

Un corollaire de l'association du concept à un terme linguistique dans le cadre de la structure inférentielle du concept est de faire ressortir le fait que cette structure inférentielle joue également un rôle dans le caractère public du concept. S'il est vrai qu'il est possible d'associer les termes "célibataire" et "être non marié" ou même les termes "célibataire" et "bachelor" ou même "unmarried man" c'est que ces termes renvoient, par hypothèse, à un *même* concept mettons le concept CÉLIBATAIRE qui apparaît dans l'espace cognitif des agents non pas sous forme langagière et linguistique mais sous la forme d'un état mental et d'un état cérébral. Et c'est du fait de la présence de cet état cognitif chez plusieurs agents que ces agents peuvent se comprendre lorsqu'ils utilisent convenablement eu égard au système utilisé (e.g. le français ou l'anglais) certains termes linguistiques. La communication entre agents cognitifs se base donc pour une part sur le concept et c'est parce que le concept est partagé, ce qu'on appelle le *caractère public* du concept, que la commu-

nication entre agents est possible. Dire que la communication entre agents est possible relativement aux concepts c'est simplement dire que si dans une discussion votre interlocuteur vous dit d'abord que Paul est célibataire et qu'ensuite dans la même discussion il parle de la femme de Paul, vous serez en droit de lui demander s'il parle du même Paul et si oui ce qu'il entend par "célibataire", dans ce cas la communication a échoué et il faudra la clarifier. La communication suppose que le locuteur partage d'une manière ou d'un autre la représentation du monde de l'auditeur. Imaginons que vous ne puissiez pas savoir d'aucune manière ce à quoi réfère le terme "célibataire" comment pourriez vous comprendre l'énoncé "Paul est célibataire". Supposer que vous comprenez ce que le terme "célibataire" dénote ne signifie pas nécessairement supposer que vous maîtrisez parfaitement la langue dans lequel ce terme est employé, on pourrait par exemple imaginer que l'on vous dise que le terme "nuliituuq" signifie "célibataire" en inuktitut et que vous soyez disposé à croire celui qui vous le dit, vous pourrez associer ce terme à ce concept sans pour autant que vous sachiez utiliser le terme "nuliituuq" dans une quelconque proposition en inuktitut vous pourriez quand même faire certaines inférences à partir du concept, dire par exemple que les gens qui parlent inuktitut doivent aussi avoir une certaine conception du couple. Mais si vous ne possédez pas le concept CÉLIBATAIRE alors vous ne pourrez pas comprendre le terme "célibataire" ni ce que vous dit votre interlocuteur. La communication suppose le concept et la publicité de celui-ci, c'est-à-dire le fait que le concept soit partagé par plusieurs agents cognitifs.

Évidemment, si l'accent est mis sur le fait que le concept puisse être partagé par plusieurs agents cognitifs – ce qui soit dit en passant ne suppose pas que tous les agents possédant le concept associent la même chose à des concepts et soient possibles de faire les mêmes inférences – cela suppose également que ces agents cognitifs possèdent la capacité inférentielle. La capacité inférentielle doit donc faire partie de l'appareillage cognitif des agents. D'autre part la publicité du concept n'implique pas que le locuteur sache d'avance, *a priori*, quel concept et associé à quel terme. Si on vous apprend, i.e. que vous ne le sachiez pas avant, que "nuliituuq" veut dire "célibataire" en inuktitut, en un sens vous avez enrichi votre concept CÉLIBATAIRE en associant un nouveau terme à sa dimension lexicale. Il est même envisageable que l'on apprenne non seulement des nouveaux termes lexicaux associés au concept mais aussi de nouveaux concepts même. Il est

possible que le concept FISSION NUCLÉAIRE ne soit pas inné et demande quelques explications avant de pouvoir être maîtrisé, ou même des concepts comme CONSTANT DE PLANK sont difficiles à fixer de manière catégorique et donc de savoir exactement quelle est son extension, intension et pertinence d'utilisation et donc quelles inférences peuvent vraiment être tirées à partir d'eux. Mais le fait que les concepts soient dans un réseau inférentiel permet justement l'apprentissage de nouveaux concepts. Il est possible d'expliquer ce qu'est la fission nucléaire en en donnant une définition peut être, en faisant un dessin ou en suivant un cours de physique. Le fait qu'il soit possible de comprendre un concept à partir d'autres concepts déjà maîtrisé laisse courir le risque d'une certaine circularité. C'est sur ce point d'ailleurs que porte les arguments contre l'analyticité [Quine, 1953], c'est-à-dire de considérer que les concepts n'auraient de sens *que* du fait de leur place au sein de la trame inférentielle. Mais la forme logique du concept n'est qu'une des dimensions du concept est si elle est nécessaire elle n'en est pas pour autant un élément suffisant.

Si les concepts sont reliés entre eux dans une trame inférentielle cela signifie qu'ils sont accessibles les uns à partir des autres. Le concept FISSION NUCLÉAIRE est probablement accessible à partir du concept ATOME. Évidemment s'il y avait un concept qui n'était relié à aucun autre ce serait assez embêtant pour expliquer à quelqu'un qui ne le posséderait pas ce à quoi réfère ce concept, mais d'un autre côté un tel concept aurait un statut assez particulier puisque cela signifierait qu'il ne pourrait tenir dans aucune inférence valide digne de ce nom – outre les pétitions de principe et les hypothèses *ad hoc* et donc ne pourrait finalement servir dans aucune représentation du monde. Ce concept, si un tel concept était possible, dépendrait en fait de ce qu'il représente pour pouvoir être utilisé, c'est-à-dire du mode de présentation de ses instances dans l'expérience et donc de la présentation non-conceptuelle de son contenu. Ce concept pourrait représenter son contenu, dans la mesure où l'agent cognitif pourrait reconnaître qu'il s'agit là l'une instance du même concept, mais il ne pourrait utiliser ce concept autrement, ce qui ne serait pas très utile. Il ne pourrait même pas croire qu'il fait l'expérience d'une instance de ce concept puisqu'alors cela serait déjà mettre en relation ce concept avec un certain contenu et avec le fait qu'il s'agisse bien d'un concept. Or de croire cela, c'est déjà croire par exemple que les concepts réfèrent à quelque chose, ou de dire

que les concepts respectent certaines contraintes inférentielles, et cela c'est déjà utiliser le concept dans une trame inférentielle.

Le fait que le concept soit enchâssé dans une structure inférentielle basée sur des capacités cognitives des agents, lui permet d'être public et donc de permettre une communication, un échange et un partage de croyances, de connaissances et de représentations du monde entre les agents. Le fait que les agents puissent partager des croyances, des connaissances et des représentations du monde suppose que les concepts soient enchâssés dans une structure inférentielle.

Tous les agents cognitifs possédant un concept ne tireront pas les mêmes inférences à partir de ce concept. Il est certain que mon concept de FISSION NUCLÉAIRE est beaucoup moins riche que celui d'un étudiant en physique nucléaire ou que celui du récipiendaire du Prix Nobel de la discipline. Il est à parier que ce dernier peut tirer plus d'inférence à partir de ce concept que je ne peux le faire. L'étendue et l'assurance avec laquelle des inférences peuvent être tirées à partir d'un concept délimitent le domaine de compétence dans la possession du concept. Le récipiendaire du Prix Nobel est plus compétent que je le suis quant à ce concept-là. Par contre il se peut que je sois plus compétent que lui quant au concept LE SCINTILLEMENT DES FEUILLES DE L'ARBRE À TRAVERS LA FENÊTRE DE MON APPARTEMENT LE 21 OCTOBRE 2004 EN DÉBUT D'APRÈS MIDI. Il s'agit là d'un concept dont l'étendue de généralité a certes moins d'application que le concept FISSION NUCLÉAIRE mais il n'en reste pas moins un concept qui est public, si quelqu'un fait l'expérience du scintillement des feuilles de l'arbre à travers la fenêtre de mon appartement le 21 octobre 2004 en début d'après midi et qu'il représente cette expérience il aura le concept ou sinon si quelqu'un qui n'était pas présent lors de l'expérience vint à lire ces lignes il pourra se faire une idée de ce que peut représenter un tel concept et l'utiliser dans des inférences. Il pourra par exemple déduire de ce concept qu'il est probable que je fus dans mon appartement le 21 octobre 2004 en début d'après midi. Mais le Prix Nobel de Physique n'a peut-être pas, pas encore peut-être, ce concept, je suis donc plus compétent que lui pour l'utiliser.

D'autre part, le fait que le concept s'articule autour de ses contenus (cognitif et objectif) de son intension, de ses conditions d'application et de sa structure logique propres, semble indiquer qu'il ne peut y avoir de concept

privé. Certes les concepts sont potentiellement articulés de manière propre et originale dans chaque agent cognitif qui pourra lui associer des choses que les autres agents cognitifs ne pourront pas saisir, mais la forme logique au moins garantie que quiconque possède la capacité conceptuelle possède un concept qu'un autre agent est en mesure de posséder.

Le domaine de compétence inclus les croyances passées et présentes de l'agent cognitif, ses connaissances, ses désirs, ses expériences et autres qui font la spécificité de cet agent.

Une certaine *division du travail conceptuel* se met même en place à l'image de la division du travail linguistique [Putnam, 1990]. Si j'utilise le concept FISSION NUCLÉAIRE je me baserais sur le concept des spécialistes de la question plutôt que sur le mien, je puisse faire référence à ces spécialistes ou lire ouvrages et si je dois tirer des inférences à partir de ce concept je pourrais me baser sur eux, par déférence, en croyant qu'ils sont plus compétents que moi dans ce domaine. Si par exemple je dis que la fission nucléaire fait intervenir les noyaux d'atomes et donc des atomes interviennent dans un certain processus nucléaire je ne suis pas en train de dire que je vais mettre ma main au feu contre cette vérité, je dis simplement que d'après ce que je sais il y a des gens qui semblent beaucoup plus qualifiés que moi sur la question et ces gens là disent que la fission requière des noyaux d'atomes.

Le fait qu'il y ait des gens plus compétents que d'autres dans certains domaines et si les domaines de compétence font intervenir des croyances, des connaissances ou toute autre représentation du monde cela implique que certains sont plus compétents que d'autres dans la maîtrise des concepts. Mais comme ces compétences sont relatives aux inférences que l'agent peut tirer à partir de ces concepts et que ces inférences impliquent la publicité du concept, tout cela permet aux gens moins compétents d'apprendre ou d'enrichir leur concept à partir des concepts des gens plus compétent. En somme enseigner a un sens. De la contrainte inférentielle posée sur le concept pour justifier de sa possibilité de représenter quelque chose découle donc son caractère public mais aussi la possibilité de son apprentissage et de son enrichissement.

Cela n'enlève rien d'une part au fait que le contenu associé au concept ne puisse pas être privé en partie – dans son contenu cognitif – ni que tous les concepts puissent être appris. Il paraîtrait paradoxal de pouvoir

apprendre quelque chose à partir de rien, c'est la fameuse question "mais comment a-t-on appris le premier concept ?" et la dissolution de ce paradoxe passe par le fait de dire que certains concepts sont déjà implémentés dans l'appareillage cognitif de l'agent. Cette position innéiste est notamment défendue par Fodor [1998]. Une autre manière de dissoudre le paradoxe serait de dire que tous les concepts peuvent être appris et que les agents cognitifs posséderaient dans leurs capacités cognitives les moyens de *construire* des concepts notamment à partir du contenu non-conceptuel de l'expérience. Cette position plus proche des empiristes tels que Locke [1997] est soutenue par Prinz [2002] en particulier.

La structure inférentielle dans laquelle se place le concept et l'impossibilité pour celui-ci de ne pas y avoir un rôle milite en faveur d'un *holisme* [Peacocke, 1992], [Dokic, 2002]. Le holisme est la thèse selon laquelle la possession de certains concepts entraîne la possession d'autres concepts et qu'il est nécessaire de posséder certains concepts pour pouvoir en posséder d'autres. Cela se comprend assez aisément à partir de l'exemple de Paul qui est célibataire et qui n'est pas marié. Posséder le concept CÉLIBATAIRE implique de posséder le concept NE PAS ÊTRE MARIÉ. Posséder un concept ne veut pas dire posséder l'ensemble de son domaine d'application et donc pour l'agent cognitif d'être compétent pour l'ensemble de ce domaine, mais pouvoir passer d'un concept à un autre par des inférences.

Ce holisme entraîne que toute modification d'un concept, par enrichissement de son extension par exemple, est susceptible d'entraîner la modification de ses relations aux autres concepts. Admettons par exemple qu'un agent cognitif mette en relation les termes "célibataire" et "bachelor" en vertu du fait qu'ils dénotent tous deux le concept CÉLIBATAIRE et imaginons que l'agent cognitif possède la croyance "Nausicaa est célibataire" et qu'avec les rudiments d'anglais qu'il a il déduise la croyance "Nausicaa is a bachelor" et que plus tard on lui fasse remarquer que le terme "bachelor" en anglais ne s'applique qu'aux hommes et que comme Nausicaa est une femme l'énoncé "Nausicaa is a bachelor" n'est pas correct en anglais bien qu'il soit vraie que Nausicaa ne soit pas mariée. Dans ce cas l'agent cognitif modifiera les termes lexicaux associés aux concepts CÉLIBATAIRE et ÊTRE NON MARIÉ et les inférences à partir de ces concepts en disant par exemple qu'on peut passer de CÉLIBATAIRE à ÊTRE NON MARIÉ mais que dans certains cas,

notamment dans les termes lexicaux associés, il faut tenir compte du sexe de la personne à qui on applique le concept. Et si un enfant ayant entendu le terme “célibataire” sans bien comprendre exactement ce qu’il veut dire mais qu’il l’associe tout de même au terme “ne pas être marié” et l’applique à son chien parce que son chien n’est pas marié sans savoir que le mariage est une institution sociale humaine, nous pourrions le lui apprendre et à ce moment-là l’enfant corrigeront les domaines des inférences accessibles au concept CÉLIBATAIRE en le restreignant au domaine des humains vivant dans en société où le mariage est une institution.

Le fait que le concept se place dans une trame inférentielle et qu’il peut être révisé par le biais de nouvelles expériences, de nouvelles croyances, connaissances ou par l’introduction – ou la suppression – de concept soulève la question de savoir comment, sur quelle base et à partir de quand un concept peut et doit être révisé. Cela revient à répondre à la question *étant donné les nouveaux éléments dont je dispose comment dois-je réviser ma représentation du monde ?*. La révision du concept se fait *a posteriori* en fonction des nouvelles données. Si les concepts sont représentés comme une distribution de probabilité dans un espace d’hypothèse, i.e. chaque concept correspond à une valeur associée à un descripteur – une hypothèse – en fonction de la propriété instanciée par l’individu tombant sous le concept et la description associée au concept dans un domaine de représentation du monde. La révision de la représentation est moins spécifique que la révision des connaissances du monde que peut avoir un agent cognitif mais le modèle de la révision des connaissances, et donc des croyances peut être étendue à la révision des concepts.

Les concepts peuvent être considérés comme des hypothèses de description c’est-à-dire des descriptions possible du monde. À ces hypothèses peuvent être associée une probabilité dont le degré varie selon la possibilité qu’une description puisse s’appliquer à un état du monde que l’on cherche à représenter. La représentation du monde est ainsi exprimée en terme de distribution de valeur de probabilité sur des descriptions. Cette modélisation permet d’appliquer la règle de Bayes aux concepts [Cornuéjols and Miclet, 2003, 3, p. 73].

Règle de Bayes :

$$p_{\mathcal{H}}(h | \mathcal{S}) = \frac{p_{\mathcal{H}}(h)P_{\mathcal{X}}^m(\mathcal{S} | h)}{P_{\mathcal{X}}(\mathcal{S})} \quad (5.3)$$

où $p_{\mathcal{H}}$ désigne la densité de probabilité définie sur l'espace des hypothèses $h \in \mathcal{H}$, $P_{\mathcal{H}}$ mesure la probabilité des événements sur \mathcal{X} et $P_{\mathcal{X}}^m$ la mesure d'un ensemble d'apprentissage $\mathcal{S} = ((x_1, u_1), (x_2, u_2), \dots, (x_n, u_n))$. $p_{\mathcal{H}}(h)$ et $p_{\mathcal{H}}(h | \mathcal{S})$ dénotent respectivement la densité de probabilité *a priori* de l'hypothèse $h \in \mathcal{H}$ et sa densité de probabilité *a posteriori* après la prise en compte des données \mathcal{S} . $P_{\mathcal{X}}(\mathcal{S} | h)$ est la probabilité conditionnelle de l'événement \mathcal{S} si l'on suppose vraie l'état du monde correspondant à h . $P_{\mathcal{X}}(\mathcal{S})$ est la probabilité *a priori* de l'événement \mathcal{S} . [Cornujéols and Miclet, 2003, 2.4.1, p. 58]

Cela signifie que pour une représentation du monde en terme de concept telle que pour chaque concept considéré comme la description d'un état du monde une valeur est associée alors si l'on prend toutes les descriptions disponibles des états du monde, il est possible de déterminer, sur cet ensemble de valeur, la valeur associée à la description d'un état du monde, i.e. d'un concept, à partir d'un ensemble d'indices donnés – par exemple par l'expérience – en fonction de sa valeur associée *a priori* (avant l'ensemble d'indices donnés) et *a posteriori*. La valeur associée au concept dépend de l'ensemble des concepts disponibles dans la représentation du monde et le fait que par hypothèse il est vrai que lorsqu'un concept est appliqué à un état du monde il décrit bien cet état du monde dans cet espace de représentation. Si la valeur associée au concept *a priori* est différente de sa valeur associée *a posteriori* on dira qu'il y a eu *révision* du concept. La révision peut être un changement dans les inférences accessibles à partir de celui-ci, dans son intension et/ou dans son extension, elle repositionne le concept sur la trame inférentielle.

5.3 Inférence et compositionnalité

Quand les jumeaux demandèrent à quoi servaient les boutons de manchette et s'entendirent répondre, par Ammu, que c'était pour

«boutonner les manches», pareille logique linguistique, dans ce qui, jusqu'ici, leur avait paru être une langue illogique, les réjouit au plus au point. *bouton + manchette = bouton de manchette*. Seules la rigueur et la logique des mathématiques pouvaient prétendre rivaliser avec un tel phénomène.

[Arundhai Roy, *Le Dieu des Petits Riens*, Folio Gallimard, 1998, p. 81]

La compositionnalité est le fait de pouvoir construire des expressions ayant un sens à partir d'autres expressions ayant un sens.

Le fait que les concepts soient enchâssés dans une trame inférentielle permet de maîtriser un nouveau concept à partir de concepts déjà maîtrisés, c'est le principe même de l'apprentissage. Pour se faire il faut pouvoir passer d'un concept à un autre et nous avons vu que cela pouvait se faire au moyen d'inférences du fait de la structure logique du concept. Le fait que la forme logique ne conditionne qu'une partie du concept cela n'implique pas qu'un concept expliqué ou maîtrisé par les bais d'un autre se réduise nécessairement à cet autre concept mais que dans l'espace des représentations – des descriptions du monde par les concepts en tant que descripteurs – l'un vaut pour l'autre extensionnellement. Par exemple dans le français courant du début du 21^e siècle dans la région parisienne, le terme “célibataire” et le terme “homme non marié” valent l'un pour l'autre dans la mesure où les concepts CÉLIBATAIRE et HOMME NON MARIÉ sont substituables extensionnellement. Le concept HOMME NON MARIÉ est dit être un *concept complexe* ou *composé* dans la mesure où il peut être *décomposé* en partie plus petites signifiantes. On retrouve dans le concept HOMME NON MARIÉ le concept HOMME qui certes est moins spécifique que ne l'est HOMME NON MARIÉ mais celui-ci est bien inclus dans le domaine du concept HOMME, tous les hommes non mariés sont des hommes. Le concept HOMME NON MARIÉ se décompose au moins en les concepts HOMME, NON, MARIÉ. Je dis au moins puisqu'il se pourrait très bien également que ces concepts se décomposent eux-mêmes en d'autres concepts.

Si les concepts sont reliés entre eux par voie d'inférences, le concept HOMME NON MARIÉ sera accessible par la combinatoire des concepts HOMME, NON et MARIÉ. Et si les concepts CÉLIBATAIRE et HOMME NON MARIÉ sont équivalents dans un mode de description alors cela implique

qu'il sera possible de posséder le concept CÉLIBATAIRE à partir des trois concepts HOMME, NON et MARIÉ.

Il faut aussi souligner que la combinatoire des concepts se fait à partir de la trame inférentielle mais que la forme logique du concept qui lui permet de se positionner dans cette trame n'est pas suffisante au concept, que celui-ci se décrit également dans ses dimensions intensionnelles, extensionnelles et pragmatiques. Il ne faudrait pas conclure hâtivement que si le concept CÉLIBATAIRE se décompose en HOMME, NON et MARIÉ, que l'extension de NON (ou de NÉGATION peut-être – il faut se souvenir que nous parlons ici de concepts et non de termes lexicaux associés) doivent nécessairement se retrouver dans l'extension de CÉLIBATAIRE. On retrouve ce genre de problème de non concordance extensionnelle dans les taxonomies forgées à partir d'héritage de propriétés. Considérez par exemple la propriété AILE ou AVOIR DES AILES. Les insectes, les oiseaux, les chauves-souris et les avions ont des ailes. Pourtant toutes ces ailes ne proviennent pas d'un ancêtre commun, c'est-à-dire que si l'on décomposait les concepts INSECTE, OISEAU, CHAUVES-SOURIS et AVIONS en concepts complexes mettons $I + AILE$ pour INSECTE, $O + AILE$ pour OISEAU, $C + AILE$ pour CHAUVES-SOURIS et $A + AILE$ pour AVION – où I , O , C et A valent respectivement pour les propriétés suffisantes dans le cadre référentiel utilisé respectivement pour INSECTE, OISEAU, CHAUVES-SOURIS et AVIONS moins le caractère AILE, et bien ce n'est pas pour autant que l'on pourra dire que si tous ces concepts partagent le même caractère que ces concepts sont co-extensifs. Il est clair qu'une aile d'avion n'a rien à voir avec une aile d'insecte d'un point de vue extensionnel. D'un point de vue intensionnel le caractère est le même, mais cela ne suffit pas pour garantir une substitution des termes. Ce problème, où un caractère commun est partagé sans que les termes soient co-extensifs est connu sous le problème de l'homologie en cladistique [[Lecointre and LeGuyader, 2001](#)].

Cette remarque à son importance lorsqu'on cherche à comprendre quelle est l'extension d'un concept complexe. L'exemple canonique discuté dans la littérature est celui du POISSON DE COMPAGNIE (*pet fish*). Si on raisonne d'un point de vue extensionnel, c'est-à-dire sur ce à quoi renvoie le concept dans le monde, on trouvera par exemple que le concept ANIMAL

DE COMPAGNIE (*pet*) inclut les chats, les chiens et les canaris et que le concept POISSON inclut les piranhas, les coelacanthes et les poissons rouges, et si généralement quand on demande à quelqu'un qu'est-ce que représente typiquement. le concept POISSON il répondra probablement quelque chose comme la truite ou le saumon et quand on lui demandera ce que représente le concept ANIMAL DE COMPAGNIE il répondra certainement un caniche ou un chat angora. Le problème donc est que si l'on essaye de fusionner ces deux extensions en une seule pour avoir l'extension de ANIMAL DE COMPAGNIE on tombera sur quelque chose comme un saumon avec une laisse ou un caniche avec des nageoires et certainement pas un poisson rouge. Cela s'explique par le fait que la composition de plusieurs concepts pour former un concept complexe ne consiste pas à coller des morceaux de concepts pour en faire quelque chose d'autre. Un concept composé reste un concept i.e. il doit se conformer aux contraintes sur les concepts notamment entrer dans la trame inférentielle, décrire quelque chose et référer à quelque chose. L'extension du concept – ce à quoi il réfère, ce qu'il dénote dans le monde – n'est qu'une partie du concept – et il est même possible que certains concepts ne dénotent rien dont on puisse faire l'expérience avec les sens (peut-être est-ce le cas de nombres ou de la conjonction par exemple) qui doit restée cohérente et consistante avec l'intension du concept – la ou les propriétés associées à ce concept – et l'utilité du concept. Le concept composé étant un concept à part entière, lui sera associé une extension, une intension et un domaine de pertinence en propre. Et si le concept composé reste dans le domaine des concepts qui le composent cela ne signifie pas pour autant qu'il en soit dépendant. On peut donc par exemple supposer que dans le cas de POISSON DE COMPAGNIE l'intérêt de la composition à partir de POISSON et de ANIMAL DE COMPAGNIE tiennent plus de l'intension que de l'extension. Pour qu'un objet du monde puisse être dit POISSON DE COMPAGNIE il faut probablement qu'il soit un poisson, i.e. qu'il ait les caractéristiques qui font de lui un poisson et qu'il ait les caractéristiques de tenir compagnie. Par exemple on pourrait dire que pour être un poisson il faut être un vertébré inférieur vivant dans l'eau et pour être qu'un animal de compagnie se définit entre autre par le fait qu'il est dépendant quant à sa subsistance, ces caractéristiques suffisent à trouver le poisson rouge, un poisson dont le Umwelt se résume à un bocal.

La structure inférentielle permet de rendre compte en partie de la compositionnalité du concept. Elle dit que si l'on veut former un nouveau concept à partir de concepts que l'on possède déjà, c'est possible de le faire en mettant en relation inférentielle les concepts intervenant dans cette combinatoire et en déterminant ensuite la position de ce nouveau concept dans la structure, en le mettant en relation avec certains autres concepts possédés, et en déterminant son extension, intension et usage.

Dans l'exemple de POISSON DE COMPAGNIE, les concepts POISSON et ANIMAL DE COMPAGNIE seront mis en relation sur la trame inférentielle, avec les propriétés des deux concepts telles qu'elles ne sont pas incompatibles – ce qui peut amener à restreindre le domaine de l'un ou des deux concepts au besoin – c'est-à-dire que si par exemple le concept ANIMAL DE COMPAGNIE est mis en relation avec le concept ÊTRE POSSÉDÉ PAR QUELQU'UN et que la propriété *être possédé par quelqu'un* n'est contradictoire avec aucune des propriétés associées au concept POISSON, alors le concept POISSON DE COMPAGNIE sera mis en rapport avec le concept ÊTRE POSSÉDÉ PAR QUELQU'UN. Par contre le concept DOIT ÊTRE EMMENÉ UNE FOIS PAS MOIS CHEZ LE COIFFEUR POUR CHIEN POUR Y ÊTRE TONDU ne sera pas associé au concept POISSON DE COMPAGNIE pour la simple raison que les poissons – même ceux de compagnie – n'ont pas de poils comme les chiens.

Le positionnement du concept complexe sur la trame inférentielle dépend donc des propriétés de celui-ci dont il hérite quelques-unes – mais pas nécessairement toutes – des concepts qui entrent dans sa construction, c'est-à-dire que son rôle inférentiel dépend de son intension. Ce n'est que dans un second temps que son extension sera déterminée. En effet, il est possible, à partir des règles d'inférences et du respect de la non-contradiction de construire à peu près n'importe quel concept. Cette construction ne dépend que des concepts que possède l'agent cognitif et non pas de ce que ces concepts peuvent représenter. Cette construction de concept est à rapprocher de la construction des idées abstraites telle qu'on le trouve dans les textes de [Locke \[1997\]](#), [Berkeley \[1988\]](#) ou [Hume \[1983\]](#). Par exemple :

Quand nous pensons à une montagne d'or, nous joignons seulement deux idées compatibles, *or* et *montagne*, que nous connaissons auparavant. [[Hume, 1983](#), II, p. 65]

Le concept MONTAGNE D'OR ne représente rien dont nous puissions faire l'expérience du fait que dans notre monde il n'y a rien de tel que des

montagnes d'or. Bien sûr rien n'empêche que dans un état différent du monde ou dans un monde possible il y a quelque chose comme une montagne d'or. Mais en fait l'extension – l'objet dont nous pourrions faire l'expérience, l'occurrence du concept – importe peu ici, ce qui compte c'est que l'*idée abstraite*, ce que nous appelons aujourd'hui un concept, est *possible* ou même serait-on tenté de dire *plausible*. Rien n'interdit de combiner les concepts MONTAGNE et OR puisqu'aucune des propriétés de l'un n'est incompatible avec les propriétés de l'autre. C'est seulement une fois le réquisit de la compatibilité, de la non-contradiction, vérifié que le problème de l'extension se pose. La compositionnalité ou la combinatoire des concepts est donc rendue possible par la forme logique des concepts, c'est-à-dire par le fait que les concepts sont reliés entre eux par biais d'inférences.

La non-contradiction dans le fait que le concept doit pouvoir représenter quelque chose semble pourtant poser une contrainte sur la combinatoire possible des concepts. Dire que le concept représente quelque chose c'est dire qu'il existe un modèle dans lequel un objet vérifie, dans un domaine donné, ce concept. Cela ne veut pas nécessairement dire que le concept doit être instancié dans le monde actuel de l'expérience, mais qu'il est possible qu'un état du monde le vérifie, que ce soit un état de ce monde-là ou d'un monde possible. Le concept DAHU par exemple représente quelque chose qui n'est pas instancié dans ce monde-ci mais qui pourrait très bien être instancié dans un monde possible, un monde où il y aurait des dahus. Le concept π est instancié dans le domaine des objets mathématiques – la question ensuite de savoir quel statut ontologique on accorde à ce domaine c'est une autre histoire.

Mais est-ce que tout ce qui est concevable détermine nécessairement un concept ? Il semble qu'il y ait des objets impossibles, comme les cercles carrés par exemple auxquels il semble qu'on ne puisse associer aucun contenu. Les objets impossibles par définition comme les cercles carrés qui incluent dans leurs compositions des concepts mutuellement exclusifs – il est possible de donner comme analyse de CERCLE le concept complexe QUI N'EST PAS CARRÉ et inversement – et par conséquent la dénotation du concept composé CERCLE CARRÉ semble s'annihiler d'elle-même. CERCLE CARRÉ pourrait être

utilisé dans une inférence, on pourrait par exemple dire : “on ne peut dessiner un cercle carré” en dérivant cette proposition d’une inférence du type :

1. Ne peuvent se dessiner que les objets possibles ;
 2. Les cercles carrés ne sont pas des objets possibles,
-
3. ∴ Les cercles carrés ne peuvent pas être dessinés.

Le problème est qu’il est tout à fait possible de construire le concept CERCLE CARRÉ à partir des concepts CERCLE et CARRÉ. Et tout composé de concepts devait donner un concept à la fin. CERCLE CARRÉ répond donc au réquisit de la forme logique du concept, i.e. de sa structure inférentielle, au réquisit de l’intension mais pas au réquisit de l’extension. C’est d’ailleurs pour cette raison que les objets impossibles sont considérés comme des objets abstraits [Zalta, 1997]. Mais il y a une différence sans doute entre le concept \wedge (la conjonction logique) et CERCLE CARRÉ. Le premier ne représente pas grand chose et il est difficile de dire ce que représente la conjonction logique en propre, mais il n’est pas impossible d’en montrer des exemples tout les mêmes ou de postuler un monde possible dans lequel existeraient en tant qu’objet logique la conjonction (peut être le troisième royaume de Frege) alors qu’il est impossible qu’il y ait aucun objet quel qu’il soit qui puisse être un cercle carré.

Faut-il donc considérer qu’il y a plusieurs types de concepts, des concepts qui représenteraient quelque chose et des concepts qui ne représenteraient rien, ou bien qu’il y a différents types d’objets que le concept peut viser et parmi eux il y a des objets qui ont la particularité de ne pas pouvoir être instancié d’aucune manière que ce soit, ou bien faut-il dire que les concepts tels que CERCLE CARRÉ ne sont pas en fait des concepts mais d’autres entités mentales issues de l’imagination par exemple ? Cette question survient ici dans le cadre de la forme logique du concept dans la structure inférentielle qui permet nous l’avons vu la combinatoire des concepts et en particulier compositionnalité et nous avons vu que l’intension du concept était utilisée en premier dans la structure inférentielle, préférentiellement à l’extension du concept. Mais que le concept déterminé sur la trame inférentielle devait répondre lui-même aux réquisits du concept. Il répond naturellement à la forme logique permettant les inférences et à l’intension puisqu’il en est di-

rectement issue, le contenu qui lui est associé – qu’il soit objectif ou cognitif – c’est-à-dire son extension ne suit pas pour autant. La question donc est de savoir si toute combinatoire sur les concepts est suffisante pour construire des concepts ou bien s’il faut évaluer indépendamment cette combinaison pour en déterminer ce qui est conceptuel de ce qui ne l’est pas. Cela découle du fait que si l’on tient le concept pour être le constituant de la pensée et en particulier de la forme propositionnelle de celle-ci, et que la pensée détermine le contenu sémantique de la forme propositionnelle, que serait le contenu d’une expression ne pouvant pas avoir – par stipulation même – de contenu associé comme dans le cas d’objet impossible ?

On pourrait objecter que la proposition “un cercle carré ne peut être dessiné” à bel et bien un contenu, elle parle de quelque chose qui ne peut exister, certes mais alors c’est un jugement sur le contenu que l’on formule, pas une saisie du contenu lui-même.

5.4 Inférence et théorie des concepts

La structure inférentielle telle qu’elle est présentée ici détermine des contraintes sur les concepts mais reste neutre quant à savoir ce qu’ils sont ou quelle conception des concepts on endosse. Fodor et al. [1980] soutient que la structure inférentielle – qui on l’a vu sous-tend une forme de holisme – est issue d’une théorie dite “classique” du concept, ce qui appelle “l’image standard” du concept (*The Standard Picture*) qui assimile le concept à une définition. La définition d’un terme détermine son extension dans la mesure où la définition énonce de manière explicite ce qu’un objet doit avoir pour que le terme – ici le concept – s’y applique. Si le concept CÉLIBATAIRE est définie par HOMME NON MARIÉ cela signifie que :

1. Tout célibataire est un homme,
 2. Tout célibataire est non marié,
-
3. ∴ Prenez n’importe quoi si c’est un homme et qu’il n’est pas marié alors vous tenez un célibataire.

Dire cela c’est dire que l’ensemble des célibataires et l’ensemble des hommes non mariés sont co-extensifs en vetu de la définition même de

CÉLIBATAIRE. La critique que dresse alors Fodor et al. [1980, p. 265] est que si c'est le cas que CÉLIBATAIRE peut être analysé en HOMME NON MARIÉ c'est du fait de cette co-extensivité. Co-extensivité que la définition de CÉLIBATAIRE par HOMME NON MARIÉ était justement supposé rendre compte. Donc quelle que soit l'extension que l'on donne à HOMME NON MARIÉ, par définition, CÉLIBATAIRE aura la même. Or d'une part la définition d'un terme – que ce soit un mot ou n'importe quel terme dans un système interprétatif – dépend de ce terme et non pas de ce qui s'y rattache. Il se peut parfaitement que l'on apprenne qu'en fait, contrairement au terme anglais 'bachelor' le terme 'célibataire' s'applique aux femmes. On révisera l'association stricte entre "célibataire" et "homme non marié", on va donc changer l'extension de l'un sans changer l'extension de l'autre sans pour autant couper la relation entre les deux termes. On pourra expliquer à un agent cognitif qui ne maîtriserait pas le terme "célibataire" à l'aide du terme "homme non marié". Donc l'extension peut changer sans que change la définition. Fodor et al. [1980, p. 265] critique également cette conception du concept qui tend à mettre en rapport inférentiels les concepts puisque selon lui la définition fixe l'extension des concepts complexes à partir de concepts de base. La question alors est celle de savoir comment l'extension de ces concepts de base est fixée.

Il est vrai que lorsqu'on entend dire que quelqu'un qui est célibataire n'est pas marié, on sous-entend que celui à qui l'on explique cela comprend que ce NON MARIÉ signifie et suppose que cet agent comprend ce que NON MARIÉ signifie en vertu du fait qu'il sait ce que NON et MARIÉ veulent dire. On peut se demander comment il le sait. Et il est fort à parier que si l'on cherche ainsi à savoir ce que l'agent doit maîtriser pour posséder un concept on entre dans un cercle peu vertueux d'un diallèle. Soit l'agent cognitif possède le concept grâce à une définition mais il faut qu'il comprenne la définition et donc qu'il possède au moins un concept, ou bien il possède les concepts de bases par un autre moyen que la définition et donc la définition n'explique pas la possession du concept.

Cet argument s'attaque indirectement à la structure inférentielle telle qu'elle est présentée ici. En effet si les concepts sont déterminés par leurs relations les uns aux autres, il faut bien qu'il y ait un commencement à cette trame et donc des concepts ou au moins un concept qui au départ

était possédé sans aucune relation à aucun autre. Et si on soutient que cette structure implique une forme de holisme alors il faudrait considérer que le premier concept est déjà un concept dans une trame holistique ce qui ne peut être le cas. Donc la trame est ouverte d'un côté ou d'un autre sur des concepts dont peut-être cette trame dépend mais dont ils peuvent se passer. En fait on peut répondre à cette objection en rappelant que si la structure inférentielle du concept est essentielle à celui-ci pour son application elle n'est pas suffisante pour déterminer le concept. La critique que formule [Fodor et al. \[1980\]](#) concerne la définition mais le fait que le concept soit explicable et analysable par d'autres concepts auxquels il est relié par des inférences ne suffit pas à dire ce à quoi réfère ce concept. Il faut le déterminer dans ses autres dimensions, que ce soit son intension, son extension et sa pertinence. Nous avons vu que nous pouvons rendre compte de CÉLIBATAIRE par HOMME NON MARIÉ même si ces deux termes ne sont pas co-extensifs. Cette explication peut néanmoins tout à fait être valable dans la mesure où elle permet, dans un contexte de saisir ce concept. Si par exemple nous sommes dans une situation où nous discutons de Paul et disons que Paul est célibataire à quelqu'un qui ne maîtrise pas ce concept, de dire que quelqu'un est célibataire si c'est un homme non marié sera peut être faux dans l'absolu mais pertinent et vrai dans ce contexte là. L'extension n'est donc pas déterminée de manière absolue mais dans le cadre d'un contexte.

Si donc les différentes dimensions du concept sont analysées de manières distinctes les unes des autres, rien n'empêche de dire que l'on peut saisir un concept à partir d'autres concepts sans que nécessairement les concepts doivent être saisis de cette manière. Il est possible d'imaginer – comme le font les empiristes par exemple – que certains concepts soient issus directement de l'expérience. Dans ce cas ils seront saisis à partir d'autre chose qu'un ou des concepts. Rien n'empêche alors d'avoir plusieurs concepts de cette manière et qu'ensuite seulement ces concepts sont interconnectés entre eux et forment une trame inférentielle qui elle-même permet d'apprendre et de construire de nouveaux concepts. Donc même si la structure inférentielle n'est pas une propriété suffisante pour construire et déterminer le concept elle n'en reste pas moins une propriété nécessaire à son application, au côté des autres dimensions du concept.

D'autant plus que si les inférences sont nécessaires à la publicité, à l'apprentissage ou à l'usage épistémique du concept, toute théorie des concepts digne de ce son se doit d'en rendre compte. Si toutes les théories des concepts n'en rendent pas toutes compte de la même manière il n'en reste pas moins que la structure inférentielle du concept est neutre quant à la théorie choisie.

On peut même se demander si certains concepts n'ont de raison d'être qu'en fonction de cette trame inférentielle. Pensez aux concepts logiques comme la conjonction par exemple. Peacocke [1992, 1.2, p. 6] rend compte du concept CONJUNCTION à partir de la capacité inférentielle :

Conjunction is that concept C to possess which a thinker must find transitions that are instances of the following forms primitively compelling, and must do so because they are these forms :

$$\frac{p}{q} \quad \frac{pCq}{p} \quad \frac{pCq}{q}$$

¹ [Peacocke, 1992, 1.2, p. 6]

Outre le fait que les trois formes de la conjonction que donne Peacocke ici ne sont pas nécessaires puisqu'il y a d'autres manières de rendre compte de la conjonction, par exemple à l'aide d'une table de vérité :

TAB. 5.1 – Table de vérité de la conjonction

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	0

¹“ La conjonction est ce concept C que pour posséder un penseur doit trouver les transitions qui sont des instances des formes primitivement irréfutables, et ce parce qu'elles sont ces formes :

$$\frac{p}{q} \quad \frac{pCq}{p} \quad \frac{pCq}{q}$$

.”

Il est également difficile de dire si formes de la conjonction données par Peacocke sont suffisantes. En effet, dans la citation ci-dessus l'agent cognitif est dit posséder le concept CONJUNCTION s'il est capable d'établir la validité des inférences en question – la première correspondant à la règle d'introduction de la conjonction, les deux secondes de sa règle d'élimination – à travers des instances de celles-ci. C'est-à-dire que si un agent cognitif se trouve face à une situation dans laquelle apparaissent un dahu et une licorne, l'agent doit être capable de dire qu'il y a un dahu dans la situation. Peacocke [1992, 1.2, p. 6] nous dit que cet agent pourra faire cette élimination de la conjonction du fait des règles qui sont irréfutables de manière primitive (*primitively compelling*). Si tel est effectivement le cas, il faut alors comprendre que l'agent cognitif possède de manière *primitive*, c'est-à-dire avant ou du moins indépendamment la possession du concept CONJUNCTION, ces règles. Il faut rappeler ici pour éviter toute méprise que la manière dont sont rendues ces règles ne soit qu'une modélisation de celles-ci, cela ne signifie pas qu'elles aient effectivement cette forme dans la cognition de l'agent, mais que cette manière de les formuler permet de rendre compte de celles que possède l'agent. Si donc nous suivons Peacocke, pour que l'agent puisse posséder le concept CONJUNCTION il faut que l'agent en question ait la capacité de faire des inférences, au moins les inférences qui lui permettent de posséder ce concept. La possession du concept CONJUNCTION au moins donc s'appuie sur une faculté cognitive autre que la faculté conceptuelle, à savoir la *capacité inférentielle*.

Il est possible de donner une liste de règles d'introduction et d'élimination pour l'ensemble des connecteurs logiques et des quantificateurs qui permettent de traduire l'ensemble des formes propositionnelles pour les calculs

des 1^{er} et 2^e ordres :

$$\wedge I) \quad A, B \vdash A \wedge B \quad (5.4)$$

$$\wedge E) \quad A \wedge B \vdash A \quad A \wedge B \vdash B \quad (5.5)$$

$$\vee I) \quad A \vdash A \vee B \quad B \vdash A \vee B \quad (5.6)$$

$$\vee E) \quad A \vee B, [A]C, [B]C \vdash C \quad (5.7)$$

$$\rightarrow I) \quad [A]B \vdash A \rightarrow B \quad (5.8)$$

$$\rightarrow E) \quad A, A \rightarrow B \vdash B \quad (5.9)$$

$$\forall I) \quad A \vdash \forall x A_t^x \quad (5.10)$$

$$\forall E) \quad \forall x A \vdash A_t^x \quad (5.11)$$

$$\exists I) \quad A_t^x \vdash \exists x A \quad (5.12)$$

$$\exists E) \quad \exists A, [A_t^x]B \vdash B \quad (5.13)$$

$$\perp I) \quad \perp \vdash A \quad (5.14)$$

$$\perp C) \quad [\neg A] \perp \vdash A \quad (5.15)$$

$$\forall^2 I) \quad A \vdash \forall X^n A_{X^n}^{P^n} \quad (5.16)$$

$$\forall^2 E) \quad \forall X^n A \vdash A_{T^n}^{X^n} \quad (5.17)$$

$$\exists^2 I) \quad A_{T^n}^{X^n} \vdash \exists X^n A \quad (5.18)$$

$$\exists^2 E) \quad \exists X^n A, [A_{P^n}^{X^n}]B \vdash B, \quad (5.19)$$

$$\lambda I) \quad S_{t_1, t_2, \dots, t_n}^{x_1, x_2, \dots, x_n} A \vdash \{\lambda x_1, x_2, \dots, x_n A\} t_1, t_2, \dots, t_n \quad (5.20)$$

$$\lambda E) \quad \{\lambda x_1, x_2, \dots, x_n A\} t_1, t_2, \dots, t_n \vdash S_{t_1, t_2, \dots, t_n}^{x_1, x_2, \dots, x_n} A \quad (5.21)$$

Ces règles d'introduction et d'élimination déterminent les déductions qu'ils sont possibles de faire à partir des termes dont on dispose dans un certain langage. Ces règles déterminent de manière *suffisante* l'usage des connecteurs sur les termes de ce langage.

For inferring certain formulas, the introduction rule gives thus a sufficient condition that is formulated in terms of subformulas of these formulas. The elimination rule, on the other hand, is related to the corresponding introduction rule according to a certain *inversion principle* : the elimination rule is in a sense only the

inverse of the corresponding introduction rule ². [Prawitz, 1965, p.8]

Qui donc manie ces règles sait manier le connecteur pour toutes ses applications possibles dans un système dans lequel ce connecteur est opérationnel, puisque ces règles sont symétriques l'une l'autre, qui manie l'une peut déduire l'autre.

Cependant, les règles d'inférences ne sont pas suffisantes pour déterminer un langage quelconque, et donc pour le maîtriser. En effet tout langage permettant de rendre compte du raisonnement, que ce langage soit formel ou non – on parle d'ailleurs de *déduction naturelle* pour l'étude de ce type de structure de raisonnement, puis que les langage formels étudient ces formes du raisonnement modélisent les raisonnements “intuitifs” [Prawitz, 1965, p. 7] – doit certes comporter des règles – parmi celle énoncées ci-dessus pour pouvoir faire des déductions, mais également comprendre un certain nombre d'hypothèses sur lesquelles ces dites règles pourront être appliquées. Le simple énoncé de ces règles – et peut-être même leur possession ne suffit donc pas à leur *application*. L'agent cognitif doit également pouvoir comprendre ce à quoi s'applique ces règles i.e. maîtriser la syntaxe et le vocabulaire du langage. Si l'agent possède les règles d'inférences de la conjonction mais n'est pas capable de reconnaître dans l'énoncé “Dario et Nausicaa sont allés au cinéma” une instance de cette règle, l'agent ne sera pas beaucoup plus avancé, puisque ces règles peuvent être restreintes dans leur application par les hypothèses disponibles dans le système.

DEDUCTION RULES. The inference rules do not characterize a system of natural deduction completely, since it is not stated in them how assumptions are discharged, and since the use of certain inference rules are circumscribed by restrictions which are formulated in terms of what assumptions the premisses depend on. In particular, to characterize the rule of $\forall E$, $\supset I$, $\forall I$, $\exists E$, and

²“Pour inférer certaines formules, la règle d'introduction donne ainsi une condition suffisante qui est formulée en termes de sous-formules de ces formules. La règle d'élimination, d'un autre côté, est reliée à la règle d'introduction correspondante suivant un certain *principe d'inversion* : la règle d'élimination est en un sens que l'inverse de la règle d'introduction correspondante.”

the principle for indirect proof, I shall state a number of rule that I shall call *deduction rules*³. [Prawitz, 1965, pp. 22-23]

La maîtrise des hypothèses sur lesquelles il est possible d'appliquer les règles de déduction est également nécessaire. Et dans le cadre d'une étude sur les concepts, si les règles d'introduction et d'élimination de la conjonction doivent d'une manière ou d'une autre être possédées par l'agent pour qu'il lui soit reconnu la possession du concept CONJONCTION, la possibilité de l'application de ces règles à des termes et donc la reconnaissance l'occurrence de ces règles dans des raisonnements et des énoncés paraît également nécessaire.

Peacocke, dans la citation précédente (5.4), nous dit que l'agent possède le concept CONJONCTION s'il est donc capable de reconnaître la validité de certaines inférences en vertu des règles formelles qui les régissent. Cela signifie que ces inférences sont réalisées dans des instances (*instances*). Mais comment peut-on instancier une règle d'inférence ?

Dario et Nausicaa sont allés au cinéma (5.22)

est une instance de conjonction, en appliquant la règle de l'élimination de la conjonction, on obtient :

Dario est allé au cinéma (5.23)

Nausicaa est allée au cinéma (5.24)

Si un agent cognitif est capable de tirer (5.23) ou (5.24) de (5.22) alors on reconnaîtra à cet agent la possession et la maîtrise du concept CONJONCTION. Cependant, pour qu'il puisse le faire, il faut qu'il puisse saisir d'une manière ou d'un autre (5.22), (5.23) et (5.24). Si ce sont bien là des instances de conjonction, elles en sont des instances particulières qui s'appliquent à des termes particuliers, que ce soit "Dario" "Nausicaa", "aller quelque part" ou

³"RÈGLE DE DÉDUCTION. Les règles d'inférence ne caractérisent pas complètement un système de déduction naturelle, puisqu'il n'est pas établi en elles comment les hypothèses sont déchargées, et puisque l'usage de certaines règles d'inférences est circonscrit par des restrictions qui sont formulées en termes des hypothèses desquelles les prémisses dépendent. En particulier, pour caractériser la règle de $\forall E$, $\supset I$, $\forall I$, $\exists E$, et le principe pour une preuve indirecte, je devrais établir un nombre de règles que je devrais appeler *règles de déduction*."

encore “cinéma”. Et ces termes ne peuvent se comprendre que dans la mesure où ils réfèrent à quelque chose et donc ne sont saisissables qu’en vertu du fait qu’ils se rattachent à des concepts, les concepts DARIO, NAUSICAA, ALLER QUELQUE PART ou CINÉMA. Certes il n’est pas nécessaire que l’agent possède effectivement ces concepts, il pourrait par exemple ne pas comprendre ce que “cinéma” signifie ou ne pas savoir qui est “Nausicaa”, mais estimer que dans le contexte il s’agit d’un lieu et d’une personne. Cependant, au niveau de l’application de la règle de la conjonction, il faut supposer que pour que cette règle soit instanciée il faut qu’elle s’applique à des termes qui sont ou se rattachent à des concepts. Cela signifie par conséquent que les concepts peuvent être mis en rapport les uns les autres avec des règles d’inférences comme celles de la conjonction par exemple. Donc les concepts doivent souscrire à la forme inférentielle. De là Peacocke [1992, 3.4, p.91] en découlera son *holisme*, ce que nous appelons ici *réseau inférentiel*.

La relation qu’ont les concepts entre eux suivant les inférences impose des contraintes sur la *référence*. Contraintes qui permettent en fait une effectivité de la référence. En effet, il est supposé que pour que (5.23) puisse être déduit de (5.22), le terme “Dario” par exemple représente là même chose dans ces deux énoncés. Les termes “Dario” réfèrent au même individu dans (5.22) et (5.23) en vertu du fait qu’ils sont tous les deux des instances du même concept à savoir DARIO. Toutes les instances d’un même concept dénotent la même chose en contexte comparable.

1. Dario dans (5.22) est une instance de DARIO
 2. Dario dans (5.23) est une instance de DARIO
-
3. ∴ Dario dans (5.22) et Dario dans (5.23) réfèrent au même individu

Si les concepts peuvent apparaître dans une trame inférentielle holistique, alors chacune de leur occurrence doit référer à la même chose dans un même contexte en vertu du fait qu’elles renvoient toutes au même concept. La référence du concept doit donc être cohérente, au moins à l’intérieur d’un même contexte. Peacocke [1992, 6.5, p. 171] appelle énoncé cette contrainte de *cohérence référentielle* ainsi :

Let us say that a possession for a concept F and a theory of how it contributes to the determination of semantic value are jointly *referentially coherent* if they ensure the validity of inferences of the following form :

$$\frac{s \text{ is } F}{\frac{s = t}{\text{so } t \text{ is } F}}$$

4

Cette subsumption vaut aussi bien pour le concept est ses occurrences que pour les concepts subsumés eux mêmes par d'autres concepts plus généraux. Par exemple les termes "Dario" dans (5.22) et (5.23) sont tous deux subsumés par DARIO, mais également DARIO et NAUSICAA – c'est-à-dire toutes les instances de ces deux concepts – sont subsumés par le concept AVEC QUI ON PEUT RIRE.

La cohérence référentielle dépend ou détermine – selon le point de vue adopté – la cohérence inférentielle, mais quoi qu'il arrive si un concept manque de satisfaire à l'inférence soit par rapport aux autres concepts soit à ses occurrences, de toute évidence ce concept sera révisé. Cela peut venir du fait que l'on découvre qu'il faut réviser son extension – dans le cas où certaines occurrences du concept entrent en conflit avec d'autres dans un même contexte ou bien si elles viennent à compromettre certaines inférences essentielles au concept – ou bien à réviser son domaine d'application – le domaine des inférences disponibles au concept.

Il faut noter que ces contraintes d'inférence et de non-contradiction auxquelles le concept doit se conformer ne dépendent aucunement des instances et donc du contenu du concept mais uniquement de sa structure formelle. C'est là un point important qui permet nous l'avons déjà indiqué plus avant de permettre de créer de nouveaux concepts dont la détermination de la référence est problématique, comme dans le cas du cercle carré par exemple.

⁴Supposons que une possession pour le concept F et une théorie de comment elle contribue à la détermination de la valeur sémantique sont *référentiellement cohérente* ensemble si elles assurent la validité des inférences de la forme suivante :

$$\frac{s \text{ is } F}{\frac{s = t}{\text{so } t \text{ is } F}}$$

.”

La trame inférentielle est un outil logique puissant, elle permet en effet, en ne s'intéressant qu'aux relations possibles entre les termes et avec pour simple contrainte la cohérence, de constituer un système expressif très grand, et ce, indépendamment de ce qu'il représente. L'élimination du contenu – nous verrons qu'il est nécessaire de prendre en compte le contenu, ou plus exactement les contenus (objectif et cognitif) du concept – correspond à l'élimination des variables dans un système formel. La réduction d'un système formel à des processus de calcul permet de considérer l'ensemble des opérations possibles au sein d'un système de manière absolue, i.e. sans devoir tenir compte du contenu particulier d'aucun des termes traité. Évidemment un tel système est pars trop réducteur et ne tient pas compte de la spécificité représentationnelle des termes, spécificité qui oblige non seulement à considérer le contenu mais également le contexte d'apparition et d'usage du terme. Mais les inférences se plient à des règles et non à des représentations. L'inférence :

Si les dahus sont boiteux alors les montagnes sont des montagnes (5.25)

est vraie en vertu d'une règle formelle et non du fait de ce que représentent les termes qui y apparaissent. En un sens on pourrait réduire 5.25 à au schéma propositionnel :

$$\varphi \supset \psi \quad (5.26)$$

Dans lequel ne compte en définitive que la relation \supset qui s'applique à n'importe quoi du moment que ce à quoi ce schéma peut s'appliquer, i.e. tout ce qui apparaît comme terme dans un langage. Cette réduction des formes inférentielles à des schémas dans lesquels disparaissent les variables remonte à Schönfinkel [1976] et à la *logique combinatoire* que l'on retrouvera dans le *lambda calcul*. En effet d'une part la logique combinatoire permet de faire ressortir la trame inférentielle et la relation entre les *blocs* du système utilisé, et le lambda calcul pourvoi des termes délestés de leurs occurrences particulières. Et finalement ce sont bien de tels termes qui apparaissent dans les schémas et même les formes propositionnelles d'inférences telles que 5.25 ou 5.22. On pourrait dire que 5.22 se comprend de la manière suivant : quoi que ce soit si c'est DARIO et quoi qu'il en soit si c'est NAUSICAA alors ce qui est DARIO et ce qui est NAUSICAA sont en relation avec quoi que ce soit si c'est

CINÉMA de telle sorte que ce qui est DARIO et ce qui est NAUSICAA sont en relation ALLER À avec ce qui est CINÉMA. Bien entendu Dario et Nausicaa sont des individus particuliers qui sont allés avoir un film particulier dans un cinéma particulier, certes, mais cela peut se comprendre que dans la mesure où le cinéma particulier possède la propriété d’être un cinéma, propriété partagée par plusieurs lieux possibles à Paris.

La décomposition des termes particuliers en porteurs de propriétés, permet de considérer les termes apparaissant dans les schémas inférentiels comme des substrats de propriétés. Par exemple si Dario et Nausicaa sont allés au cinéma “Grand Action”, “Grand Action” portera la propriété d’être un cinéma, il ne sera qu’une occurrence possible de cette propriété, or pour comprendre que le “Grand Action” est un cinéma il faut être en mesure de le subsumer sous la propriété *cinéma* en question. Cette propriété étant commune à tout cinéma tout en ne pouvant être saisie indépendamment d’un particulier, on pourra l’extraire, l’abstraire de “Grand Action” de la sorte :

$$\text{Cinéma(Grand Action)} \quad (5.27)$$

$$\lambda(\text{Grand Action Cinéma})[(\text{Grand Action})] \quad (5.28)$$

ou pour plus de lisibilité, en utilisant des variables et des constantes logiques nous aurons pour la forme schématique de 5.28 :

$$\lambda\varphi \quad (5.29)$$

dont on peut déduire φ qui vaut pour la propriété qui caractérise l’objet qui la porte. C’est grâce à cette propriété que l’on peut reconnaître, attribuer et appliquer à un objet que l’on peut isoler, individuer et identifier un objet du monde. Cette propriété utilisable dans le schéma inférentiel n’étant pas instanciée à ce stade – il reste encore à déterminer sa référence, celle justement que l’on vient de lui enlever – elle sera dit être abstrait [Prior, 2002, p. 56]. La propriété détermine ce qu’il y a de commun à toutes les instances relevant du même concept. En ce sens le concept agit comme un ensemble dont ses occurrences sont les membres. La propriété étant le label étiqueté sur l’ensemble. C’est cette manière de voir le concept qui amène à le considérer en terme ensembliste et à faire du concept un prédicat attribué à ses instances [Bealer, 1982, Peacocke, 1992]. La propriété isolée, c’est-à-dire détachée de l’occurrence qui la supporte permet d’assigner une

connotation au terme apparaissant dans les schémas inférentiels et donc de distinguer les termes les uns des autres.

L'usage des propriétés pour distinguer les termes revient à réduire ceux-ci à des définitions. Par exemple on dira que l'objet o_1 est un chat parce qu'il possède toutes les propriétés nécessaires et suffisantes pour être CHAT. Autrement dit on réduit le concept CHAT à une liste de propriétés nécessaires et suffisantes qui permettent de subsumer un objet les satisfaisant sous ce concept. Si par exemple un chat est un "petit mammifère familier à poil doux, aux yeux oblongs et brillants, à oreille triangulaires et griffes rétractiles, qui est un animal de compagnie" [selon le dictionnaire *Le Petit Robert*, 1996], alors le concept CHAT sera associé à cette liste de caractères, et un objet pourra être un chat uniquement s'il vérifie ces caractères. Cette conception du concept est défendue parce qu'il est convenu d'appeler la *théorie classique des concepts* (cf. chapitre 3.1).

Cette conception qui vise à réduire les termes utilisés dans des inférences à des définitions se retrouve dans le programme constructiviste de Hilbert and Bernays [2001a,b]. En effet, dans ce programme les concepts mathématiques sont construits à partir d'opérations et de définition, de sorte qu'à partir d'un nombre restreint d'opérations et de termes il est possible de construire et de déduire l'ensemble des termes pouvant apparaître dans le système. Tout terme construit à partir de cette axiomatique compte comme pouvant être utilisé dans le système pour construire ou déduire d'autres termes à partir des opérations disponibles, de fait, tout ce qui importe et que les termes puissent être utilisés dans le système et non ce qu'ils représentent ou dénotent. En prenant ce programme constructiviste comme modèle pour comprendre les concepts, on remarque que c'est de la même manière que l'on justifie des concepts comme DAHU, MONTAGNE D'OR ou CERCLE CARRÉ. Quoi que si l'on contraint le système sur les propriétés avec un principe de cohérence, des concepts contradictoires quant à leur intension tel que CERCLE CARRÉ seraient écartés de la trame conceptuelle. Si par exemple CERCLE est associé avec la propriété complexe : "courbe plane fermée dont tous les points sont à égale distance R " [*Le Petit Robert*, 1996] et le CARRÉ est associé avec la propriété complexe : "qui forme un quadrilatère dont les angles sont droits et les quatre côtés égaux", alors aucun objet ne pourra avoir la conjonction des propriétés de CERCLE et de CARRÉ. En un sens le système pourra donc diviser son domaine en *possible* et *impossible*. Le

problème qui surgit alors est celui du statut qu'il faut conférer à cette distinction. En effet, le possible et l'impossible ici concernent l'*application* du terme à un objet. Or nous avons vu que du point de vue de la définition sans un système d'opération, la question de l'extension ne se pose pas. Elle ne se pose pas parce qu'alors cela signifierait que l'on part de l'objet visé par la définition pour en déduire la définition alors qu'on agit exactement dans l'ordre inverse. Sinon en effet nous ne pourrions jamais parler de dahu puisque personne n'en a jamais fait l'expérience et ce du fait même que dans la définition du dahu il est mentionné que l'on ne peut en faire l'expérience. Le concept en tant que *fonction caractéristique* n'est donc pour le moment que déterminé par la place qu'il occupe dans la trame inférentielle, et comme le concept a pour fonction de viser des objets du monde, il est supposé que le concept représente quelque chose sans que nous puissions pour autant déterminer son extension. Et évidemment la simple non-contradiction dans une structure inférentielle n'est pas suffisante pour dire si l'extension est effective, i.e. réalisée dans le monde de l'expérience ou non. Tout ce qu'on peut dire c'est que le concept ainsi déterminé *peut* représenter quelque chose, mais ce n'est pas nécessaire qu'il soit instancié. Cette possibilité permet déjà tout de même de dire que le concept représente alors quelque chose.

Le principe de représentation est une propriété importante des termes de la trame inférentielle, et c'est sans doute ce fait qui encourage à penser le concept comme un prédicat. Tout comme l'unicité de la représentation est une propriété intéressante d'un terme, suffisamment du moins pour qu'on la signale par un opérateur spécifique, l'opérateur γ [Russell, 1989, pp. 268-269], la représentation d'un terme ou d'une formule – une chaîne de termes composés – également intéressante à spécifier, c'est ce que fait l'opérateur ϵ de Hilbert [Hilbert and Bernays, 2001b, II, §1] :

Le ϵ -symbole constitue par conséquent une sorte de généralisation du μ -symbole pour un domaine d'individus quelconque. D'après sa forme, il représente une fonction d'un prédicat variable qui, en dehors de l'argument auquel se rapporte la variable liée relative au ϵ -symbole, peut encore contenir des variables libres comme arguments ("paramètres"). La valeur de cette fonction, pour un prédicat \mathfrak{A} déterminé (après fixation des paramètres) est une chose du domaine d'individus; et plus précisément, cette chose, conformément à la traduction

matérielle de la formule (ϵ_0) , est *une chose à laquelle ce prédicat \mathfrak{A} s'applique, étant supposé qu'il s'applique à au moins une chose du domaine d'individus*. [Hilbert and Bernays, 2001b, II, §1, p. 62]

En somme le symbole ϵ indique l'extension d'un terme construit à partir d'un réseau inférentiel. C'est ce que feront les quantificateurs dans le calcul des prédicats en délimitant le domaine d'application d'une propriété entendue alors comme un prédicat. Mais il est *supposé* qu'une propriété puisse avoir un domaine d'application, ce domaine n'est pas nécessairement donné. C'est pour cela que l'on peut supposer que le concept DAHU représente bien quelque chose dans le sens où il vise les objets qui sont des dahus – les porteurs de la dahuité – où que ces objets se trouvent, et s'il n'y a pas d'objets comme ça dans le monde actuel de l'expérience le fait même que ce concept est bien construit permet de postuler qu'il existe un monde possible où il pourrait y avoir des objets porteurs de la propriété associée au concept DAHU. L'application de l'opérateur ϵ au concept DAHU indique simplement que ce concept s'applique à quelque chose en vertu même du fait qu'il est un concept et qu'un concept représente quelque chose.

L'opérateur ϵ agit donc comme parallèle à l'opérateur λ . Alors que l'opérateur λ indique que seule l'intension d'un terme est prise en compte, l'opérateur ϵ indique que l'extension est considérée. Cette distinction demande encore à être clarifiée mais nous voyons se pointer déjà la structure du concept, la détermination d'un terme dans une trame inférentielle auquel est associée à la fois une propriété – une intension – et un domaine d'objets auxquels il s'applique – une extension. Alors qu'un terme préfixé du λ sera dit être un *abstrait* [Church, 1956, §03], un terme préfixé par un ϵ serait dit être un *idéal* [Avigad and Zach] dans la mesure où l'opérateur ϵ permet de dire que toute formule libre de tout quantificateur est dérivable alors elle l'est également dans le calcul des prédicats, en traduisant les quantificateurs en termes préfixés de l'opérateur ϵ . La définition ϵ_0 (5.30) [Hilbert and Bernays, 2001b, II, §1, p. 63] permet d'éliminer les quantificateurs :

$$\exists x(A(x)) \rightarrow A(\epsilon_x A(x)) \quad (5.30)$$

ou

$$(A(a)) \rightarrow A(\epsilon_x A(x)) \quad (5.31)$$

à l'inverse la formule (5.32) permet de les introduire :

$$A(\epsilon_x A(x)) \rightarrow \exists x(A(x)) \quad (5.32)$$

Ce qui établit l'équivalence entre les formules quantifiées et les formules préfixées de l'opérateur ϵ , ce qui donne a définition ϵ_1 (5.33) [Hilbert and Bernays, 2001b, II, §1, p. 64] :

$$\exists x(A(x)) \equiv A(\epsilon_x A(x)) \quad (5.33)$$

L'extension des termes des donc idéalisée afin de réduire les opérations sur ses termes à des *schémas d'inférences* – dans le cadre du programme des fondements des mathématiques à une trame déductive de preuve [Hilbert and Bernays, 2001b, II, §1, p. 70].

L'opérateur ϵ en attribuant une extension idéale à une propriété permet ainsi donc de trancher les concepts dont la référence ne peut pas être atteinte dans le monde actuel. L'attribution d'une extension idéale fait donc sortir la question de la référence du cadre du possible et de l'actuel. Tout terme, en tant qu'il est bien construit se voit d'office attribuer une extension en tant que ce terme détermine une propriété qui en tant que telle permet de décrire quelque chose. Ainsi en va-t-il pour le concept CERCLE CARRÉ, s'il peut être construit dans un système inférentiel cohérent réfère à quelque chose par stipulation. La distinction entre les termes abstraits, tels que les nombres ou les objets mathématiques par exemple, et les termes empiriques ou recognitionel, tels que chat ou rouge par exemple, n'importe donc plus puis que pour chaque terme qui apparaît dans la trame inférentielle une extension idéale peut être postulée de cette manière.

Les opérateurs d'abstraction – l'opérateur λ – et d'idéalisation – l'opérateur ϵ – seront entendus, dans le cadre d'une théorie des concepts en termes respectivement d'*intension* et d'*extension*. Bien sûr l'abstraction et l'idéalisation ne sont pas à proprement parler des fonctions déterminant

l'intension et l'extension d'un terme, mais elles permettent de le faire dans le cadre des concepts, dans la mesure où le concept s'articule autour de ces dimensions. L'abstraction correspond à la détermination de la propriété associée au concept, l'idéalisation à la détermination de sa référence.

Les opérateurs λ et ϵ sont symétriques l'un l'autre. En effet, les règles d'introduction et d'élimination de l'opérateur ϵ affirment que s'il existe un objet dans le domaine d'une forme propositionnelle, alors il est possible d'attribuer une extension idéale à cette forme propositionnelle. D'autre part si une extension idéale peut être assignée à une forme propositionnelle, alors il est possible de considérer que son domaine d'application n'est pas vide. Cela découle du fait même que c'est une forme propositionnelle, i.e. qu'un modèle la vérifie et donc que ce modèle n'est pas vide. En réécrivant les règles d'introduction et d'élimination pour l'opérateur ϵ (5.30) aux regards des règles d'introduction et d'élimination du quantificateur \exists (5.12) et (5.13), nous obtenons :

$$\epsilon I) \quad A_t^x \vdash A_{\epsilon_x} A \quad (5.34)$$

$$\epsilon E) \quad A_{\epsilon_x} A, [A_t^x] B \vdash B \quad (5.35)$$

C'est-à-dire qu'il est possible de déterminer une extension pour toute propriété qui apparaît instanciée. Si une occurrence de propriété apparaît alors il est possible de dire, en fonction de cette occurrence, que cette propriété s'applique à quelque chose sans déterminer plus avant ce domaine d'application. Au contraire, si un domaine d'application a été déterminé pour une propriété alors cela signifie que cette propriété s'applique au moins à quelque chose, i.e. cette occurrence. La détermination de l'extension en ce sens est donc *a posteriori* d'après les données que nous possédons déjà, celles-ci provenant de l'expérience.

A contrario, l'opérateur λ nous informe que pour toute forme propositionnelle s'appliquant à un domaine, il est possible de considérer qu'une propriété s'applique à chacune des occurrences dans le domaine, à savoir cette forme même, et que pour toute propriété qui s'applique à un domaine, il est possible d'instancier cette propriété dans une forme en l'appliquant à un objet du domaine. La propriété en question permet donc d'individuer et d'identifier chacun des éléments du domaine en lui attribuant une pro-

priété permettant de le faire apparaître, à savoir la propriété qu'il possède quelle que soit celle-ci. En somme, les règles d'introduction et d'élimination de l'opérateur λ peuvent se comprendre sous cette forme :

$$\lambda I) \quad \forall X^n A \vdash \lambda x A \quad (5.36)$$

$$\lambda E) \quad \lambda x A \vdash A_{T^n}^{X^n} \quad (5.37)$$

L'opérateur λ permet donc de caractériser les propriétés qui déterminent la forme propositionnelle comme propriété permettant d'*identifier* des objets du monde quels qu'ils soient s'ils satisfont une propriété. Cette détermination est donc une détermination de détermination (de 2^e ordre) affirmant simplement que si un objet peut être identifié, alors il satisfait une propriété quelconque. Nous sommes donc bien à un niveau *conceptuel* et non pas non-conceptuel, puisqu'il faut identifier l'objet et non se contenter de l'individuer. Si donc une individuation est possible, elle implique qu'il soit possible d'isoler la propriété caractérisant l'objet en question. Or si pour toute propriété, une extension est déterminable – sans pour autant supposer l'identité à ce stade, l'individuation suffisant – alors pour toute forme propositionnelle précédée par l'opérateur λ il est possible de déduire cette même forme propositionnelle préfixée de l'opérateur ϵ et qu'il est possible de préfixer d'un λ toute forme issue de l'élimination de l'opérateur ϵ . L'extension idéale désignée par l'opérateur ϵ correspond à la *référence postulée* développée au chapitre 4.

La trame inférentielle dans laquelle se situe le concept permet met donc de le comprendre par rapport aux autres concepts dont dispose l'agent cognitif qui le possède. Pour l'heure elle ne permet pas de rendre compte de ce que dénote ou de ce que connote le concept, nous verrons comment nous pouvons rendre compte de ces dimensions du concept.

L'idée d'une trame inférentielle pour comprendre un terme à travers les opérations qu'il permet au sein de cette trame, c'est-à-dire de comprendre ce terme comme un *processus* [Peacocke, 1992, 2.1, p. 46] ou comme *fonction* [Putnam, 1975, 1990, Dennett, 1997, 1993] dans un système représentationnel recoupe la *sémantique des rôles conceptuels* (*Conceptual Role Semantics*), selon laquelle la signification d'une représentation est déterminée et ex-

pliquée par le rôle de cette représentation dans l'espace cognitif de l'agent qui la possède [[Greenberg and Harman](#)].

Chapitre 6

Changement conceptuel

Lorsqu'il retira son masque, en sortant dans l'air frais du matin, il se rendit compte qu'il venait d'absorber une dose massive d'un lointain recoin de la réalité dont une grande majorité du pays ne soupçonnait même pas l'existence. Cela jeta une lumière entièrement neuve sur le concept de dinde de Noël garnissant un rayonnage de supermarché.
[Tristant Egolf, *Le seigneur des porcheries*, Folio, Gallimard, p. 218]

Un agent cognitif peut être amené à changer les concepts qu'il possède. Par exemple un agent peut apprendre que l'homme qui est accoudé au bar là bas n'est pas Paul comme il le pensait, mais que c'est Pierre. Ou bien que l'actuel président de l'autorité palestinienne n'est plus monsieur Yasser Arafat mais monsieur Mahmoud Abbas, ou bien que les baleines ne sont pas des poissons mais des mammifères, ou bien encore qu'il faut distinguer la disjonction inclusive de la disjonction exclusive lorsqu'on consulte la carte des menus au restaurant, etc.

Si un concept est une entité mentale possédée par un agent cognitif et qui permet à cet agent de représenter son environnement pour former des jugements, des croyances, des connaissances, des désirs ou des plans d'action sur lui, alors cela signifie que le concept est assimilé à entité inférée à partir d'un ensemble de raisons qui le soutiennent. Ces raisons ne sont pas nécessairement bonnes, mais ce sont celles auxquelles se réfère l'agent pour utiliser le concept tel qu'il le fait. Si l'agent tient le concept comme dérivant bien de cet ensemble de raisons – i.e. s'il tient ces raisons comme étant de bonnes raisons pour justifier son usage du concept alors cela signifie que pour cet agent, l'inférence du concept à partir de ces raisons est valide et

fiable :

$$\lceil \Gamma \vdash \alpha \rceil \quad (6.1)$$

et donc :

$$\alpha \in \Gamma \quad (6.2)$$

Si l'agent au contraire pense que le concept n'est pas soutenu par des *raisons*, que ces raisons lui laissent penser que ce concept ne s'applique pas dans ce cas, alors il pourra dériver non application de ce concept :

$$\lceil \Gamma \not\vdash \alpha \rceil \quad (6.3)$$

et donc :

$$\alpha \notin \Gamma \quad (6.4)$$

Dans les cas où l'agent suspend son jugement, c'est-à-dire où il ne se prononce pas de manière catégorique, cela signifie que l'inférence du concept à partir des données qu'il possède ne lui permettent pas d'assigner et d'asserter le concept avec assurance. Autrement dit l'agent considère qu'il se pourrait en fait que ce concept ne dérive pas de ces raisons. Cela peut se comprendre de deux manières : ou bien l'agent hésite entre l'application de ce concept et l'application de sa négation, ou bien il hésite entre l'application de ce concept et l'application d'un autre concept, ces deux concepts n'étant pas nécessairement contradictoires – l'un peut être plus précis que l'autre – ou bien être peuvent être exclusifs mais relever de raisons compatibles. Imaginez les cas où l'agent ne sache pas si l'objet qui lui est présenté est un chien ou un loup. Il juge que les indices dont il dispose ne lui permettent pas de trancher de manière catégorique entre les deux concepts. Ces deux concepts sont contradictoires au moins sur le plan métaphysique. Si quelque chose est un chien alors ce n'est pas un loup et inversement. L'agent pourrait décider de trancher l'affaire en disant que l'objet en question est un animal, le concept ANIMAL subsumant les concepts CHIEN et LOUP – i.e. CHIEN et LOUP se superposent suivant le caractère ANIMAL – à ce compte là, l'agent

n'aura pas assez d'indices à son avis pour dériver ou bien CHIEN ou bien LOUP mais suffisamment pour dériver ANIMAL.

Est-ce à dire que l'agent doit toujours être en mesure de dériver la disjonction du concept ou de sa négation ? C'est-à-dire qu'un agent devrait au moins croire en l'application disjonctive du concept ?

$$(PADC) \quad \lceil \Gamma \vdash \alpha \vee \neg \alpha \rceil \quad (6.5)$$

Mais ce *principe d'application disjonctive du concept* PADC est trop fort. En effet cela suppose que l'agent accepte le principe du tiers-exclu, or le fait qu'il suspende quelque fois son jugement, laisse penser que dans certains cas au moins, l'agent ne sait pas s'il est pertinent d'assigner ce concept ou non à l'objet en question. Cela vient du fait non pas qu'un agent puisse nécessairement croire tout et son contraire, cela irait à l'encontre du *principe de cohérence* exigé par la rationalité attendue de l'agent PACR, mais au fait que l'agent ne sache pas nécessairement tout, i.e. qu'il ne sache pas dériver l'ensemble des conséquences logiquement accessibles des raisons qu'il invoque ou bien qu'il ne connaisse pas nécessairement l'ensemble des raisons qui soutiennent un concept. Il faut donc relativiser le principe d'application disjonctive aux concepts et raisons possédés par l'agent, et non pas le considérer dans l'absolu au sens de dérivations dans un système logique complet possédé par un agent idéal.

Le fait qu'un agent ne puisse pas dériver ou bien α ou bien $\neg\alpha$ de Γ n'implique pas nécessairement une violation du principe de rationalité PACR, du moment que l'agent ne dérive pas d'une manière qu'il pense fiable et justifiée $\alpha \wedge \neg\alpha$ de Γ , puisqu'alors cet agent violerait la consistance de son inférence, et dirait n'importe quoi.

Il s'agit donc non seulement de tenir compte des concepts que possède un agent mais également des justifications qu'il utilise pour utiliser ces concepts et de la manière dont il justifie ces concepts à partir de ces raisons. Un agent peut être amené à changer ses concepts, au moins certains. La référence d'un concept peut changer et ne plus s'appliquer au même objet – c'est le cas par exemple du concept PROFESSEUR À PARIS 4, dont l'extension peut varier au cours du temps –, ou bien les justifications qui soutiennent un concept peuvent changer – la limite d'âge pour postuler

au CNRS peut être modifiée du jour au lendemain et changer les raisons d'application du concept CONDITION D'APPLICATION AU CNRS – ou même peut être les règles d'inférence d'un concept à partir d'un ensemble de raisons peuvent être modifiées.

6.1 Raisons pour de changer de concept

Un agent peut être amené à changer de concept dans cinq cas paradigmatiques : l'arrivée de nouveaux indices qui remettent en cause le concept même, deux concepts co-extentionnels et/ou co-intensionnels sont jugés contradictoires, l'une des raisons utilisées pour justifier l'usage d'un concept s'avère être fausse, l'extension du concept change sans affecter son intension, de nouveaux éléments dans les raisons ou les applications sèment le doute sur l'usage actuel du concept.

6.1.1 De la non pertinence avérée du concept

La première raison qui engage l'agent à réviser l'un de ses concepts, est la non pertinence avérée de celui-ci. L'agent utilisait un concept selon certaines raisons, et de nouveaux éléments plaident en faveur du fait que ce concept n'est plus pertinent ni dans son usage ni dans son application. C'est sans doute ce qui s'est passé avec un concept tel que PHLOGISTIQUE. Certains auteurs tels que J. J. Becher ou G. E. Stahl ont pu développer le concept de PHLOGISTIQUE – en fait le construire de toute pièce – afin de rendre compte et d'expliquer le phénomène de combustion. Les propriétés affectées à ce concept sont des plus étranges puisqu'il s'agit d'une substance incolores, inodore, insipide, sans poids et qui disparaît durant la combustion, il est donc difficile d'appliquer positivement ce concept à aucun objet, son application se fait dans le cadre d'une explication scientifique, afin de rendre compte d'un phénomène mal compris en assignant toutes les propriétés inexplicables à un élément nouveau. Les recherches chimiques s'étant par la suite affinées, des auteurs tels que A. L. Lavoisier ou C. L. Berthollet ont montré que l'on pouvait se passer du concept obscur de PHLOGISTIQUE et expliquer le phénomène de combustion à l'aide des concepts OXYDATION et RÉDUCTION. Ainsi le concept PHLOGISTIQUE a perdu son extension, il ne désigne plus une substance mais un phénomène – on pourrait penser que le terme désigne ne

fait l'oxygène, mais toute combustion ne requière pas nécessairement de l'oxygène, l'hydrogène pouvant également faire l'affaire – et ses conditions d'usages puisqu'il s'agit en fait d'une conjonction de deux phénomènes – l'oxydation et la réduction – indépendant l'un de l'autre. Le concept de phlogistique devient donc vide extensionnellement et intensionnellement, ce qui amène à l'abandonner. Évidemment il est toujours possible de dire que le concept PHLOGISTIQUE garde toujours une extension et une pertinence d'usage, mais cela uniquement maintenant d'un point de vue historique – e.g. pour comprendre les travaux physique de la fin du XVII^e il faut supposer et comprendre que ce concept a une certaine pertinence. En un sens donc le concept n'est pas perdu, il est simplement si fortement amendé qu'il est relégué aux archives et n'est plus utilisé, et se voit remplacé par d'autres concepts tels que OXYDATION et RÉDUCTION. Il s'agit là donc d'un radical, pour ne pas parler tout simplement d'abandon – et d'extension et d'intension – du concept, parce que celui-ci a perdu toute pertinence d'usage.

Si donc :

$$\lceil \Gamma \vdash \alpha \rceil \quad (6.6)$$

et que

$$\Gamma \rightarrow \Gamma^* \quad (6.7)$$

tel que

$$\Gamma^* \doteq \Gamma \wedge \xi \quad (6.8)$$

et que

$$\Gamma \wedge \xi \not\vdash \alpha \quad (6.9)$$

alors ou bien l'agent doit abandonner α s'il pense que $\Gamma^* \vdash \neg\alpha$ ou bien remplacer α par α^* s'il pense que $\Gamma^* \vdash \alpha^*$.

6.1.2 Deux concepts coréférentiels différents entraînent la révision ou l'abandon de l'un des deux au moins.

Si un agent possède les concepts α et β – par exemple EVEREST et SAGARMATHA – et qu'il pense que $\Gamma \vdash \alpha$, $\Delta \vdash \beta$ mais pense que $\Gamma \not\vdash \beta$ et $\Delta \not\vdash \alpha$, alors l'agent tient α et β pour des concepts différents, i.e. comme ayant des conditions d'usage ou des conditions d'applications différentes.

De fait deux concepts sont tenus pour être différents s'ils diffèrent dans leurs conditions d'usage, c'est-à-dire que les raisons qui justifient l'usage de l'un ne justifient pas l'usage de l'autre, ce qui correspond à la distinction entre les concepts proposée par Peacocke :

Distinctness of Concepts : Concepts C and D are distinct if and only if there are two complete propositional contents that differ at most in that one contains C substituted in one or more places for D , and one of which is potentially informative while the other is not ¹. [Peacocke, 1992, p. 2]

Cela signifie donc que les deux concepts ne sont pas substituables l'un l'autre dans un jugement.

À la différence des raisons, s'ajoute la différence de référence, il est possible – au moins théoriquement – d'imaginer que deux concepts α et β soient soutenues par les mêmes raisons mais qu'ils diffèrent de référence, c'est-à-dire que α s'applique à l'objet o_1 alors que β s'applique à l'objet o_2 mais que α ne s'applique pas à l'objet o_2 et β à l'objet o_1 . En fait ce cas théorique est très étrange, puisque les conditions d'application sont généralement données dans les raisons justifiant l'usage du concept – un agent justifiera son usage du concept α dans son application à o_2 par le fait que l'objet en question possède une certaine propriété ou caractéristique P et l'une des raisons d'usage de α spécifie que lorsque P est présente dans un objet alors cet objet relève de α . La distinction extensionnelle de concepts co-intentionnels est d'autant plus étrange que – comme nous l'avons vu – le concept peut se réduire à un terme d'abstraction à la Takeuti, i.e. à une

¹“*La distinction des concepts* : les concepts C et D sont distincts si et seulement si il y a deux contenus propositionnels complets qui diffèrent au plus dans le fait que l'un contient C substitué dans une ou plusieurs places pour D , et que l'un est potentiellement informatif alors que l'autre ne l'est pas..”

fonction caractéristique [Church, 1956, Prior, 2002] ne tenant pas compte de l'extension du concept. Si donc les concepts ne sont pas distingués sur leur fonction caractéristique alors ces deux concepts sont les mêmes, et ils ne peuvent donc pas différer quant à leur extension. Prenez même le cas d'une fonction caractéristique qui concernerait la structure chimique d'un élément et en spécifierait son agencement pour désigner un objet. Il se peut alors parfaitement que cette caractéristique permet de désigner sans possibilité de les distinguer un diamant et un morceau de carbone de charbon de bois, alors même que d'un agent distinguerait sans doute un diamant du charbon de bois. mais alors, il en ressort que la co-intentionnalité qui conduit à des extensions que l'agent considérerait comme différentes, résulte en fait d'une insuffisance de spécification intensionnelle. Il est probable par exemple que ce même agent subsumerait sous le concept CAILLOUX un diamant et un morceau de charbon, mais cela ne l'empêcherait pas de pouvoir spécifier les différences extensionnelle qu'il fait entre les deux, par des raisons qui s'appliqueraient alors non plus au concept CAILLOU mais aux concepts – qu'il tiendrait pour distincts – de DIAMANT et CARBONNE, la différence entre les deux ne serait peut être pas la structure micro-chimique, mais la valeur sociale des objets désignés par ces concepts. Ce qui signifie donc que si les extensions de deux concepts sont distinguées par un agent cognitif, leurs termes d'abstraction le sont aussi, et donc il n'est pas possible d'avoir deux termes strictement co-intensionnels avec des extensions différenciées.

Imaginons donc qu'un agent possède deux concepts qu'il tient pour différents selon leurs conditions d'utilisation et d'application. En d'absence de toute autre information, l'agent va utiliser le *principe de référence postulée*, et comme selon lui les raisons qui justifient l'usage du premier concept – mettons EVEREST – sont différentes que celles qui garantissent le second – mettons SAGARMATHA – l'agent va postuler deux extensions différentes pour ces deux concepts. L'agent peut par exemple penser que EVEREST et SAGARMATHA désignent des montagnes et se dire, suivant la maxime de la redondance de la communication – en fait le rasoir d'Ockham – selon laquelle il est inutile de donner deux noms à une même chose sans justification raisonnable, qu'il doit s'agir de deux montagnes différentes. Maintenant imaginons que cet agent planifie de gravir l'ensemble des 8000 de la planète et pour cela compulse des cartes topographiques et

découvre en fait que ces deux concepts désignent en fait une seule et même montagne, mais que son nom – et les propriétés qui lui sont conférées – change d’un versant à l’autre. L’agent va probablement réviser ses concepts EVEREST et SAGARMATHA (et peut-être également ceux CHOMOLUNGMA et QOMOLANGMA qui sont les autres noms donnés au Mont Everest) en ajoutant dans les conditions d’usage et les raisons qui soutiennent ces deux concepts, le fait qu’ils soient co-extensionnels dans la mesure où ils réfèrent au même objet dans le monde. Que ces concepts soient co-extensionnels ne signifie pas nécessairement qu’ils soient co-intensionnels, leurs conditions d’usage et même d’application peuvent différer. En effet le fait d’utiliser le terme ‘Everest’ pour désigner la plus haut montagne de la planète, implique que l’on utilise la désignation occidentale de cette montagne – cette montagne ayant été baptisée ‘Mont Everest’ l’honneur du chef du Service géodésique de l’Empire britannique des Indes. Si cette même montagne est appelée ‘Chomolungma’ par les tibétains et que ce nom signifie “déesse-mère” dans cette langue, et si cette même montagne est appelée ‘Sagarmatha’ par les népalais et que ce nom signifie “déesse du ciel” dans cette langue, les concepts EVEREST, CHOMOLUNGMA et SAGARMATHA ne sont pas identiques intensionnellement. Un agent se tromperait pas exemple s’il disait que EVEREST signifie la déesse du ciel pour les népalais, et ce même s’il est vrai que le Mont Everest en tant que c’est la montagne la plus haut de la planète, désigne bien la déesse du ciel pour ce peuple.

Une fois que l’agent apprend la co-extension de deux concepts, il peut ou bien décider l’éliminer l’un des concepts si les intensions de ces concepts s’avèrent – pour lui au moins – être identiques :

$$\begin{array}{l|l} 1 & \epsilon\alpha \dot{\equiv} \epsilon\beta \\ 2 & \epsilon(\alpha, \beta) \\ 3 & \ulcorner \alpha \dot{\equiv} \beta \urcorner \end{array}$$

De même si les concepts sont co-intensionnels, l’agent – s’il se conforme au principe de rationalité [PARC](#) – devra ternir ces concepts comme iden-

tiques :

$$\begin{array}{l|l}
 1 & \lceil \Gamma \dot{\vdash} \alpha \rceil \\
 2 & \lceil \Delta \dot{\vdash} \beta \rceil \\
 3 & \lceil \Gamma \dot{\equiv} \Delta \rceil \\
 \hline
 4 & \lceil \Gamma \dot{\vdash} \alpha, \beta \rceil \\
 5 & \lceil \Delta \dot{\vdash} \beta, \alpha \rceil \\
 6 & \lceil \alpha \dot{\equiv} \beta \rceil
 \end{array}$$

Si l'agent considérait les concepts α et β comme différents et vient ensuite à les considérer comme étant identiques, i.e. co-extensionnels ou co-intensionnels, l'agent – s'il se conforme au principe de rationalité **PARC** – doit réviser ses concepts α et β en α^* et β^* incluant l'équipotence entre ces concepts ou bien sur l'intension – α^* et β^* ont les mêmes conditions d'usage – ou bien sur l'extension – α^* et β^* ont les mêmes conditions d'application – ou bien sur les deux.

Il est évident qu'un agent possédant les concepts α – mettons EVEREST – et β – mettons SAGARMATHA – qu'il considérerait maintenant comme co-extensifs, ne devrait pas nécessairement abandonner l'un des deux. Cet agent peut toujours penser que ces deux concepts diffèrent intensionnellement – que l'un est pertinent dans un contexte de discussion sur les croyances népalaises alors que l'autre l'est moins ou pas du tout. Ces concepts seront traités par cet agent comme synonymes et substituables l'un l'autre d'un point de vue extensionnel mais pas nécessairement du point de vue intensionnel. Il n'empêche qu'une fois la co-extension apprise par l'agent, ses concepts sont révisés pour l'introduire dans les raisons qui les justifient : le concept α supporté par les raisons Γ , deviendra α^* :

$$\alpha^* := \lceil \Gamma \vee \Delta \dot{\vdash} \alpha \rceil \quad (6.10)$$

Si l'agent considère en fait que les raisons pour soutenir α et β sont en fait les mêmes – i.e. qu'il n'y a aucune différence intensionnelle entre ces deux concepts – et que α et β sont co-extensifs, alors l'agent, par respect du rasoir d'Ockham – abandonnera l'un des concepts au profit de l'autre, ce

qui revient à réduire l'un de ces concept à l'autre, sorte de révision radicale du concept qui s'apparente au premier cas de révision.

6.1.3 L'affaiblissement ou le renforcement des raisons garantissant l'usage du concept entraîne la révision de celui-ci

Une troisième raison amenant un agent à réviser un de ses concepts est l'affaiblissement des conditions d'utilisations de celui-ci. Imaginons un agent apprenant le concept DAHU – concept dont il ignorait tout jusqu'alors – de la part de quelqu'un qu'il croit digne de confiance, i.e. auquel il défère pour l'usage de ce terme. Imaginons que cet agent crédule croit que DAHU désigne réellement un animal dans le monde, puisque rien ne lui a été dit sur l'existence ou la non existence de ces animaux et qu'il en déduit charitablement, que ce dont on parle sans en mentionner explicitement la non-existence, existe. Quelque temps après, cet agent apprend qu'il a été l'objet d'une farce et qu'en fait les dahus n'existent pas. L'agent va réviser son concept de DAHU en ajoutant la restriction, dans les raisons justifiant son usage, en spécifiant que le concept DAHU s'applique aux objets ayant les caractéristiques P plus la particularité de ne pas exister dans le monde.

$$\lceil \Gamma \vdash \alpha \rceil \quad (6.11)$$

sera affaibli en :

$$\lceil \Gamma \wedge \xi \vdash \alpha^* \rceil \quad (6.12)$$

Cette révision peut avoir de sérieuses conséquences sur le concept α . D'une part son extension peut-être considérablement révisée, si l'agent pensait qu'un animal jusque-là inconnu de lui et observé en zone montagneuse était tenu par lui pour être un dahu, l'agent, une fois qu'il aura appris la non-existence des dahus dans le monde de l'expérience – par conséquent dans la région de montagne où il a vu le chamois qu'il prenait pour un dahu – revoir *a posteriori* l'application de ce concept à cet objet.

L'extension devra être changée si le nouvel indice qui affaiblit les raisons justifiant l'usage du concept est maintenant contradictoire avec la référence postulée pour le concept par l'agent.

$$\begin{array}{l|l}
1 & \lceil \Gamma \dot{\vdash} \epsilon \alpha \rceil \\
2 & \lceil \Gamma \wedge \xi \dot{\vdash} \epsilon \alpha \rceil \\
\hline
3 & \lceil \Gamma \wedge \xi \dot{\vdash} \epsilon \alpha^* \rceil
\end{array}$$

Une version radicale de l'affaire serait que le nouvel indice contredise l'ensemble des raisons qui soutiennent le concept, et donc amèneraient à abandonner complètement celui-ci :

$$\begin{array}{l|l}
1 & \lceil \Gamma \dot{\vdash} \alpha \rceil \\
2 & \lceil \Gamma \wedge \neg \Gamma \dot{\vdash} \alpha \rceil \\
\hline
3 & \lceil \Gamma \wedge \neg \Gamma \dot{\vdash} \alpha \wedge \neg \alpha \rceil \\
4 & \alpha \doteq \perp
\end{array}$$

De manière symétrique, le renforcement des raisons entraîne une révision du concept :

$$\begin{array}{l|l}
1 & \lceil \Gamma \dot{\vdash} \epsilon \alpha \rceil \\
2 & \lceil \Gamma \vee \xi \dot{\vdash} \epsilon \alpha \wedge \epsilon \alpha_{+\xi} \rceil \\
\hline
3 & \lceil \Gamma \vee \xi \dot{\vdash} \epsilon \alpha^* \rceil
\end{array}$$

Si Γ est l'ensemble des raisons qui permettent de dériver le concept α et si la contraction de Γ avec ξ permet de dériver à la fois le concept α et le concept β qui n'est autre que $\alpha + \xi$, alors le concept α doit être révisé pour lui ajouter ξ dans ses caractéristiques.

L'affaiblissement des raisons provoque en fait l'élargissement de l'extension – de l'application – du concept.

Un agent peut par exemple croire pour une raison ou une autre que les cygnes sont blancs et apprendre qu'il existe des cygnes noirs, et modifier les conditions d'usage du concept en enlevant la spécification de la couleur pour

l'application du concept CYGNE à un objet. L'élimination d'une contraire dans la caractérisation de l'objet pour lui assigner un concept a pour effet de renforcer les caractéristiques du concept qui déterminent son application. Ce renforcement élargi le champ d'application du concept et donc son extension.

6.1.4 La modification de l'extension du concept modifie celui-ci

Parallèlement à la modification intensionnelle, l'extension du concept peut conduire à sa révision : un nouveau objet dans le monde auquel s'applique un concept peut entraîner la modification de ses caractéristiques. Imaginons qu'un agent possède le concept L'ACTUEL PRÉSIDENT DE L'UKRAINE et qu'il pense, pour une raison ou une autre que ce concept ne peut s'appliquer à un et un seul objet. Imaginons maintenant qu'à la suite d'élections pas tout-à-fait transparentes, deux candidats différents revendiquent le statut de président de l'Ukraine. La référence du concept L'ACTUEL PRÉSIDENT DE L'UKRAINE s'en voit bouleversée, puisqu'alors qu'elle ne comprenait par définition qu'un seul objet, voilà qu'elle désigne deux objets dans le même monde. L'agent peut décider de changer son concept L'ACTUEL PRÉSIDENT DE L'UKRAINE en changeant ses conditions d'application – et le cas revient au cas précédent – et dire qu'en fait ce concept n'est pas une description définie mais que plusieurs objets peuvent la vérifier en même temps, ou bien il peut décider que l'extension doit à tout prix être préservée comme unique et affaiblir son application et distinguant maintenant deux concepts : L'ACTUEL PRÉSIDENT DÉMOCRATIQUEMENT ÉLU DE L'UKRAINE et L'ACTUEL PRÉSIDENT NON DÉMOCRATIQUEMENT ÉLU DE L'UKRAINE, chacun deux étant une description définie et s'appliquant à l'un des deux objets du monde.

Le cas peut se présenter sous la forme d'un changement d'extension sans changement intensionnel. Par exemple, lorsqu'il a été découvert que les baleines n'étaient pas des poissons mais des mammifères, l'extension des concepts POISSON et MAMMIFÈRE s'en est trouvée modifiée (par exemple le mammifère le plus gros n'était plus l'éléphant d'Afrique mais la baleine bleue). L'intension des concepts POISSON et MAMMIFÈRE ne change pas, seule la description des objets du monde qui change, c'est-à-dire l'application de ces concepts à l'environnement.

$$\begin{array}{l|l}
1 & \lceil \Gamma \dot{\vdash} \epsilon\alpha \rceil \\
2 & \lceil \Gamma \dot{\vdash} \epsilon\alpha \wedge \neg \epsilon\alpha_{\xi} \rceil \\
\hline
3 & \lceil \Gamma \dot{\vdash} \epsilon\alpha^* \rceil
\end{array}$$

où $\epsilon\alpha^*$ correspond à $\epsilon\alpha_{\cdot\xi}$.

Mais ce type de changement est très étrange. En effet, l'hypothèse de la prédominance de l'intension contredit le fait qu'il puisse y avoir un changement au niveau de l'extension qui n'ait aucune répercussion au niveau intensionnel, à moins qu'il ne s'agisse pas d'un changement de concept en tant que tel, mais simplement un enrichissement de l'environnement d'éléments déjà catégorisés par ailleurs. En somme rien de change dans la représentation conceptuelle de l'environnement.

Le cas extrême de changement extensionnel qui aurait des répercussions au niveau intensionnel serait celui d'un ensemble vide qui viendrait à comprendre maintenant un élément. Mais ce changement ne serait pas qu'extensionnel, puisqu'il impliquerait que la propriété "s'appliquer à quelque chose" ou pour le dire de manière plus radicale, le fait d'"exister" viendrait à apparaître dans la fonction caractéristique du concept. Cependant, comme par le *principe de référence postulée* (cognitive ou canonique) tout concept possède une référence au moins idéale, par essence, aucun concept n'est vide. La notion de "concept vide" est donc contradictoire. De fait ce type de changement purement extensionnel n'est pas possible. Par conséquence, l'affirmation selon laquelle l'intension ne change pas, seule la description change dans ce cas là est mise à mal, et aucune description ne peut changer sans impliquer une modification au niveau intensionnel.

6.1.5 Toute suspicion sur l'usage ou l'application du concept doit conduire à sa révision

Tout indice – extensionnel ou intensionnel – soulevant le doute sur l'application ou l'usage d'un concept doit entraîner la révision de celui-ci ou au moins la suspension de son utilisation par l'agent. En fait il ne s'agit pas tant ici d'une révision à proprement parler que du respect d'une certaine mesure d'utilisation du concept. Imaginons qu'un agent possède un concept α et qu'il pense l'utiliser à bon escient mais que plus il apprend des choses à son propos, moins il sait exactement quelles sont exactement ses

conditions d'utilisation et d'application – cela peut être le cas de concepts métaphysiques par exemple. L'agent, s'il veut se conformer au principe de rationalité devrait se méfier de ce concept et éviter de l'utiliser tant qu'il n'est pas absolument certain de pouvoir le faire de manière appropriée puisqu'en définitive il est difficile de dire à quoi s'applique réellement le concept – quelle est son extension – et même quelles sont les justifications de son usage. L'agent devrait donc, pour se conformer au principe de rationalité amender ce concept du fait qu'il est un concept incertain et obscur, et essayer ou bien de le remplacer par des concepts plus assurés, précis et déterminés ou bien essayer de déterminer, si cela est possible, de manière plus précise se concept en révisant son extension ou son intension. Ce type de révision se ramène donc aux cas précédents. Sa particularité est peut être que contrairement au cas précédents il met l'accent sur l'usage pragmatique du concept. L'agent ne doit pas utiliser ses concepts n'importe comment. Si le concept est une entité mentale qui permet à l'agent qui le possède de représenter son environnement, cela signifie que le concept doit avoir un rôle et une utilité "avérés" dans cette représentation. Si cette représentation est inutilisable par l'agent parce qu'il est incapable de savoir et de comprendre ce que ce concept représente réellement, le bon sens de la raison voudrait que l'agent abandonne ce concept en l'état, car finalement il ne fait que surcharger l'espace cognitif de l'agent sans y apporter de contribution positive. Rien évidemment n'oblige ni structurellement – i.e. les règles d'inférences du concept à partir de ses justifications – ni du point de vue contenu – l'extension et les règles d'application du concept – l'agent à abandonner ou réviser ce concept. Ici il s'agit plutôt d'une règle de conduite du bon usage du concept et de l'espace cognitif, en somme il s'agit d'appliquer le rasoir d'Ockham au entités mentales et éviter de les multiplier à l'excès plus qu'il n'est besoin. L'agent devrait laisser de côté les concepts qui lui paraissent obscures et dont l'usage, la fonction et l'utilité lui sont douteuse ou insaisissable, ou bien il devrait les réviser jusqu'à que le doute ne les entoure plus. Mais il s'agit là d'un principe quasi éthique de l'usage du concept, et même si est souhaitable d'exiger que l'agent se conforme au principe de rationalité, rien ne peut le forcer à la servitude rationnelle.

6.2 Réquisits d'un changement de concept

L'agent cognitif peut changer de concept pour les différentes raisons énoncées dans la section précédente. Un agent cognitif se pliant au *principe d'agent cognitif rationnel*

$$(PACR) \quad B \diamond (\ulcorner \epsilon \alpha \urcorner_{a_i} \not\equiv \ulcorner \epsilon \alpha \urcorner_C)$$

changera son concept si celui-ci est en désaccord avec le concept canonique et ne permet pas d'établir une communication avec un autre agent. Si l'agent pense que son concept est en désaccord avec le concept canonique, et s'il pense qu'un expert le corrige ou bien que les informations qui lui sont données lui permettent de réviser son concept alors il est attendu de cet agent qu'il révisé son concept. Cette attente est une contrainte sur l'usage que l'agent fait de ses concepts, l'agent s'y plie uniquement s'il veut jouer le jeu de la communication. Cette révision est donc normative et se fonde sur l'usage social des concepts. Cela ne signifie pas que tout concept doit nécessairement être social et partagé, mais simplement que le jeu de la communication et du partage des représentations du monde suppose un terrain d'entente supposée. Si un agent s'accroche à ses concepts sans jamais vouloir les réviser et que ceux-ci sont en conflit avec les concepts canoniques leur correspondant, il est fort à parier que cet agent sera mis en marge de la société comme excentrique ou fou. L'agent est libre de cet usage privé des concepts, et ses concepts restent bien des concepts au sens défini ici, i.e. des entités mentales qui permettent à l'agent qui les possède de représenter son environnement pour le comprendre ou agir dessus. Seulement cet agent vivra dans 'son' monde sans tenir compte – et donc – sans reconnaître les autres agents et leurs concepts. Donc cet agent représentera bien le monde, mais sera jugé autiste et asocial. Ce point est intéressant à soulever, puisqu'il signifie que le principe de publicité, invoqué pour la justification de la structure holistique et épistémique du concept, n'est peut être pas nécessaire. Il est possible de penser une version purement privée du concept et de rattraper le phénomène social – le fait que les agents communiquent et se comprennent lorsqu'ils partagent des représentations du monde – par un phénomène purement évolutionnaire. Les concepts pourraient être des entités mentales dépendantes de l'évolution de l'espèce humaine – au moins – entités qui représenteraient l'environnement de

manière purement téléosémantique, en ce sens le concept serait purement inné [Fodor, 1998] et le contenu et l'usage du concept ne dépendrait ni des actions, des attentes, de la construction du monde par l'agent, mais celles-ci dépendraient uniquement des concepts dont disposerait l'agent, et si deux agents parviennent à communiquer et à s'entendre sur la manière dont ils représentent et comprennent le monde cela serait dû au fait qu'ils appartiennent tous deux à la même espèce au même stade de l'évolution et donc partagent le même appareillage cognitif et donc les mêmes concepts. Et comme les concepts représentent ce qu'ils représentent indépendamment de leur porteur, tous les agents représentent le monde de la même manière. Cela suppose que les concepts soient écrits dans un langage commun à l'ensemble des agents de la même espèce partageant le même espace cognitif. Ce langage est le langage de la pensée, le mentalais [Fodor, 1974].

Le problème avec cette approche est qu'elle ne permet pas d'expliquer l'erreur d'usage ou d'application du concept par un agent. En effet, si la représentation du concept ne dépend pas, en aucun sens de l'histoire de l'agent, alors un agent ne se trompe jamais lorsqu'il utilise un concept pour représenter quelque chose. Car ou bien ce concept ne s'applique pas à cet objet mais alors l'agent ne peut pas le savoir, ou bien l'agent sait à quoi le concept s'applique et l'applique volontairement à mauvais escient mais alors cela signifie que l'agent peut d'abord se représenter la chose par le concept approprié et ensuite choisir un concept non approprié pour en parler, mais il ne s'agit plus alors d'application ou d'usage de concept, mais de l'élaboration d'un jugement épistémique de très haut niveau. En somme l'agent ne se trompe pas mais ment – “se trompe intentionnellement” – ce qui n'est pas tout à fait la même chose. Or l'intérêt dans la possession d'un concept est d'expliquer comment un agent peut changer, au cours de son histoire, de représentation du monde et donc de concept.

Il y a des concepts que possède un agent cognitif qu'il est difficile de lui attribuer de manière innée. Le concept `CONNEXION BLUETOOTH` par exemple est certainement un concept que je ne possédais pas il y a quelques années encore. Évidemment reste la stratégie de dire que ce concept est en fait rien d'autre qu'un composé de concepts que je possédais déjà. Certes, mais alors ou bien tous les concepts se réduisent à un certain nombre de

concepts primitifs acquis de manière innée ou bien certains concepts non primitifs contiennent de l'information qui n'est pas réductible à la conjonction de concepts primitifs. Dans le premier cas le seul apprentissage envisageable est l'apprentissage de liaisons entre concepts. Cette liaison n'apportant rien de nouveau puisque l'ensemble de ses composants sont déjà connus et maîtrisés par l'agent. Quand à la liaison elle-même elle provient du langage de la pensée et donc est elle-même déjà possédée par l'agent. Dans ce cas, un nouvel élément du monde produit par l'espèce à laquelle appartient l'agent sera sans doute réduit sans trop de problème à une représentation issue de concepts primitifs innés. Par exemple une connexion Bluetooth sera simplement "une connexion à distance par ondes entre deux appareils", i.e. le concept CONNEXION BLUETOOTH se réduira au concept complexe – i.e. composé de concepts simples agencés entre eux par des connecteurs mentaux – CONNEXION À DISTANCE PAR ONDES ENTRE DEUX APPAREILS, et si certains de ces concepts ne sont pas innés, alors ils seront remplacés par le concept complexe composé de concepts primitifs.

Mais qu'en est-il si un nouvel objet apparaît dans l'environnement de l'agent sans que son apparition provienne d'aucune construction de la part de l'espèce à laquelle appartient ? Par exemple qu'en est-il des premiers contacts avec ornithorynques ? Jusqu'en 1798 l'agencement et la classification des êtres vivants dans la nature semblait relativement stable et établie et voilà qu'apparaît un mammifère ovipare – un animal qui pond des œufs et allaite ses petits – animal qui bouleverse complètement la représentation des animaux. Qu'est-ce qu'un agent cognitif doit faire face à un tel animal ? Bien sûr il est possible de penser qu'un tel animal était imaginable avant même sa rencontre – les agents cognitifs vivants avant 1798 possédaient le concept d'ovipare et de mammifère et rien ne les empêchait d'imaginer un monstre répondant à la conjonction des deux. Mais il y a une différence entre postuler une référence (PRP) et confronter cette référence à un objet existant du monde, i.e. appliquer le concept. Il n'est pas très difficile de croire que les concepts MAMMIFÈRE et OVIPARE aient été modifiés après 1798. Alors qu'ils étaient supposés être exclusifs, voilà qu'ils ne l'étaient plus. Cela change leurs conditions d'usage et leurs applications, i.e. leurs intensions et leurs extensions. Peut-être que le partisan de l'innéisme dira qu'en fait les concepts complexes peuvent être modifiés parce qu'en fait ce ne sont pas vraiment des concepts, seuls les concepts primitifs et innés eux ne

changent pas. Certes mais alors cela signifie que les concepts MAMMIFÈRE et OVIPARE ne sont pas primitifs et que les combinaisons de concepts primitifs qui les composent ne sont pas bonnes, bien que leurs composants le soient. Mais cela signifie bien tout de même que ces composants innés ont été mal utilisés et mal appliqués pour que les composés en collent pas à la réalité. D'une manière ou d'une autre donc il faut pouvoir rendre compte du fait qu'une erreur est possible. Or si l'agent n'est pas à l'origine, d'une manière ou d'une autre de son application et de son usage du concept alors il n'est pas possible de rendre compte de l'erreur.

Si Paul dit que l'oiseau qui est perché dans l'arbre en face est un freux alors qu'en fait il s'agit d'un chouca, Paul se trompe. Il perçoit un oiseau mais l'identifie mal. Le fait que Paul perçoive un oiseau avec telle et telle propriété – la taille, la couleur, etc. – c'est-à-dire qu'il discrimine un objet du monde parmi d'autres objets du monde – l'oiseau de l'arbre – ne relève pas du concept, i.e. ni de son usage ni de son application, mais de l'appareillage perceptif de Paul, i.e. du niveau non-conceptuel. Le fait qu'il discrimine un objet et qu'il pense que cet objet est un freux, relève de la description, de la représentation que fait Paul de l'objet discriminé, i.e. Paul identifie l'objet en question comme étant un freux, i.e. il pense que le descripteur "freux" correspond à l'objet en question. Le descripteur "freux" correspond à un certain ensemble d'objet et pas aux objets qui n'appartiennent pas à cet ensemble du fait de certaines propriétés que les objets doivent posséder afin de faire parti de cet ensemble. Paul peut ne pas connaître l'ensemble des propriétés qu'un objet doit avoir pour être un chouca mais ne pas être un freux, et cette ignorance est responsable du fait qu'il applique le terme 'freux' à ce qui est en fait est un chouca. Et l'usage du terme 'freux' dépend de ce que Paul croit que ce terme désigne, représente dans le monde, i.e. du concept FREUX. Un ornithologue pourra corriger Paul et lui dire qu'en fait l'oiseau perché dans l'arbre n'est pas un freux mais un chouca et lui dire pourquoi. Cette explication permettra à Paul de modifier son concept FREUX et d'enrichir, voire même d'apprendre, le concept CHOUCA. L'application et l'usage du concept CHOUCA ne requiert pas nécessairement de concepts primitifs. L'ornithologue peut simplement montrer un objet dans le monde et le désigner comme étant un chouca et sur cette base Paul enrichira son espace conceptuel du concept CHOUCA. On pourra très bien dire qu'en fait Paul comprendre qu'un chouca est un oiseau parce que l'objet désigné est

un oiseau. Certes, mais Paul pourrait très bien apprendre le concept OISEAU de la sorte.

L'ostension n'est qu'une manière d'apprendre un concept. Paul pourrait très bien apprendre le concept CHOUCA dans un cours de biologie sans aucun exemplaire à l'appui, mais uniquement par comparaison par exemple avec ce que sont un freux et une corneille, i.e. n'apprendre que les propriétés que doit avoir un objet pour être un chouca sans être ni un freux ni une corneille ni autre chose. L'empirisme n'est donc pas la seule alternative à l'innéisme. La seule chose qu'il est important de souligner ici est que la discrimination – i.e. la perception et l'individuation de propriétés qui font qu'un objet est ce qu'il est – ne dépend pas du niveau conceptuel – i.e. de l'usage et de l'application d'un concept – mais du niveau non-conceptuel. Que ce niveau dépend en parti ou complètement d'un appareillage inné est une chose, sa représentation au niveau conceptuel en est une autre.

6.2.1 Les concepts d'un agent doivent être consistants

L'agent qui change de concept doit respecter un *principe de consistance*. En effet, pour être tenu pour rationnel, un agent ne doit pas entretenir de concepts contradictoires et incohérents.

L'incohérence et l'inconsistance au niveau conceptuel peut se situer ou bien dans les conditions d'usage du concept ou bien dans ses conditions d'application. Au niveau de l'usage, cela signifie que les mêmes raisons soutiennent deux concepts différents et mutuellement incompatibles.

$$\begin{array}{l} \Gamma \vdash \alpha \neg_{a_i} \\ \Gamma \vdash \neg \alpha \neg_{a_i} \end{array}$$

Un agent possédant ces deux concepts simultanément dans son espace cognitif violera le principe de consistance, puisque à partir de ces deux concepts il pourra dériver n'importe quoi :

$$\Gamma \vdash \alpha \wedge \neg \alpha \neg_{a_i} \rightarrow \Gamma \equiv \perp \quad (6.13)$$

Il est évident qu'avec un concept de ce genre, l'agent peut tout représenter et n'importe quoi, un objet relevant et ne relevant pas en même temps de ce concept. Un tel agent ne pourrait pas rendre compte de l'usage de son concept puisque les raisons qu'il donnerait seraient les mêmes que

celles qu'il donnerait pour soutenir le non usage de ce concept dans cette situation. Pour être rationnel, un agent doit être consistant. Au niveau conceptuel cette consistance s'exprime par la cohérence des concepts possédés par un agent. Le premier réquisit de la possession d'un concept est donc le respect du principe de consistance *principe de consistance* (PC) :

(PC) L'ensemble des concepts possédés par un agent doit être cohérent.

Cela signifie qu'un agent ne doit pas posséder deux concepts co-extensionnels ou co-intensionnels contradictoires.

6.2.2 Les concepts d'un agent doivent être clos inférentiellement

Si un agent possède les raisons Γ telles que de ces raisons, le concept α peut être inféré, alors l'agent doit posséder le concept α . L'agent doit se conformer au *principe normatif de clôture inférentielle* du concept (PNCI) :

$$(PNCI) \quad \Gamma_{a_i} \vdash \alpha \rightarrow \ulcorner \Gamma \vdash \alpha \urcorner_{a_i}$$

Ce principe est normatif. Il est demandé à ce que l'agent le respecte, c'est-à-dire que lorsqu'il possède tout ce qui est requis pour posséder un concept, que cet agent possède ce concept, i.e. qu'il infère ce concept des raisons qui le justifie et que l'agent possède. Évidemment un agent peut refuser d'inférer un concept, et nul ne peut l'en empêcher. Mais un tel agent sera considéré comme étant de mauvaise foi, puisqu'il ne fera pas ce qui est en pouvoir de faire.

Toutefois ce principe ne requière pas que l'agent sache nécessairement tout ce qui peut s'inférer d'un ensemble de raison. Il n'est pas demandé à l'agent d'être omniscient. En effet sinon l'agent devrait connaître par exemple l'ensemble des tautologies logiques et être en mesure de les inférer de toute raison consistante. Ce qui est demandé à l'agent est de respecter le *principe de pertinence* selon lequel l'agent doit utiliser le concept le plus approprié pour représenter ce qu'il cherche à représenter, et de posséder le concept le plus approprié, selon l'agent – qui peut être inféré à partir des raisons possédées par l'agent en question.

Ce principe est nécessaire pour penser que l'agent se conforme au principe de rationalité cognitive PARC. En effet, pour qu'un agent puisse ac-

cepter que ses concepts puissent différer des concepts canoniques, il faut que l'agent possède des concepts. Or si un concept est une entité mentale qui représente et qui possède des condition d'usage et l'application, il faut que l'agent accepte ces conditions. Si l'agent possède les conditions en question mais ne reconnaît qu'elles puissent servir à justifier l'usage et l'application d'un concept, l'agent ne peut accepter que son concept puisse différer de celui des autres, puisqu'il ne reconnaît pas avoir le concept. Au moins minimalement – dans les limites des raisons possédées par l'agent et des contextes où se trouve l'agent, i.e. les situation d'application – l'agent doit inférer les concepts pertinents des raisons qu'il possède.

La raison d'être du principe de clôture inférentielle du concept, principe qui peut paraître trop fort réside dans le fait qu'il permet d'énoncer les principes de remises en cause positives et négatives.

6.2.3 Principe de remise en cause positive

Le principe de remise en cause positive (*Principe of Positive Undermining*), énoncé par Harman [1986] affirme qu'un agent devrait cesser de penser φ lorsque cet agent pense de manière positive que l'une ou plusieurs des raisons qu'il possède pour croire φ ne sont pas bonnes.

One should stop believing φ whenever one positively believes one's reason for believing φ are not good ². [Harman, 1986, p. 39].

Ce principe, Harman l'énoncé pour les croyances. Mais dans la théorie inférentialiste des concepts telle qu'elle est développée ici, ce principe d'applique également. Si un agent possède le concept α d'après la justification Γ et qu'en même temps il pense ou bien que $\Gamma \equiv \perp$ ou bien que $\Gamma \not\vdash \alpha$, alors l'agent ne devrait plus accepter le concept α et le réviser.

D'une manière positive, cela signifie qu'un agent doit penser que les justifications qui soutiennent ses concepts sont de bonnes justifications, c'est-à-dire sont fiables, vraies et utiles. Là encore il s'agit d'un principe normatifs, puisque l'agent peut être de mauvaise foi et continuer à utiliser un concept en sachant pertinemment que ce concept n'est pas justifier tel qu'il devrait

²“ Quelqu'un devrait arrêter de croire φ à partir du moment où il croit positivement que l'une des raisons pour croire φ n'est pas bonne.”

l'être. L'une des conséquences du non respect de ce principe est qu'un agent sachant pertinemment que son concept n'est pas en accord avec le concept canonique correspondant continue tout de même à utiliser son concept sans vouloir le réviser.

6.2.4 Principe de remise en cause négative

Symétrique au principe précédent, et toujours en suivant [Harman](#), le *principe de remise en cause négative* (*Principle of Negative Undermining*) affirme qu'un agent devrait cesser de penser φ lorsqu'il n'associe pas à φ une justification appropriée.

One should stop believing φ whenever one does not associate one's belief in φ with an adequate justification (either intrinsic or extrinsic) ³. [[Harman, 1986](#), p. 39].

Appliqué au concept cela signifie qu'un agent ne devrait pas utiliser ou appliquer un concept pour lequel il n'a pas de justification – ni bonne, mauvaise – ou bien qu'il utilise un concept sur la base d'une raison qu'il ne pense pas correspondre à ce concept.

Ce principe peut se comprendre de deux manières. Dans l'une de ses versions il correspond au principe précédent. L'agent possède une raison qui justifie le concept qu'il utilise mais cet agent sait que cette raison n'est pas bonne pour soutenir ce concept. Une autre version serait de dire que l'agent ne possède pas de raison justifiant de son usage et de son application du concept. Soit qu'il n'a absolument aucune raison et donc utilise le concept sans suivre aucune règle d'inférence ou d'application, soit qu'il ne sait pas si les raisons qu'ils utilisent sont bonnes ou pas, i.e. justifient bien ce concept ou non. Dans les deux cas, l'agent devrait au moins suspendre son jugement et réviser son concept en conséquence ou bien au plus rejeter ce concept. Quoi qu'il en soit si l'agent n'est pas capable de dire pourquoi il utilise son concept de la manière dont il le fait, l'agent peut en être blâmé, et il lui est demandé de réviser son concept.

³“Quelqu'un devrait arrêter de croire φ à partir du moment où il n'associe pas une des croyances dans φ avec une justification adéquate (qu'elle soit intrinsèque ou extrinsèque).”

De manière positive, ce principe stipule qu'un agent doit toujours être en mesure de donner les raisons qui justifient – selon lui – l'usage et l'application qu'il fait de ses concepts.

6.2.5 Principe de préférence dans le changement

Tous les concepts ne se valent pas dans l'espace conceptuel d'un agent. Les concepts dont l'agent est certain de leur usage et de leur application – i.e. l'agent croit que les raisons qui justifient ce concept sont bonnes et fiables de telle sorte qu'il est capable d'affirmer ou de nier l'usage et l'application de ce concept dans toutes ou presque les situations – n'ont pas le même poids dans l'espace cognitif de l'agent que les concepts pour lesquels il n'est pas certain des raisons qu'il donne sont bonnes et assurées, si bien qu'il préfère suspendre son jugement.

La révision des concepts doit se faire suivant l'ordre d'assurance que l'agent a dans ses concepts. La structure holistique des concepts suppose que l'agent a la possibilité de changer un concept plutôt qu'un autre pour préserver la consistance de ses représentations compte tenu de l'environnement. Par exemple, une fois confronté à l'ornithorynque, le biologiste peut choisir ou bien de refondre complètement sa classification des êtres vivants ou bien de rajouter une case dédiée à l'ornithorynque dans sa classification et en ajoutant simplement les modifications. Le *principe de préférence dans le changement* (PPC) dit simplement que l'agent doit préférer changer d'abord les concepts dont il est le moins sûr de leurs conditions d'usage et d'application et seulement ensuite et en cas de besoin, les concepts qu'il tient pour assurés et certains.

(PPC) Doivent être révisé de préférence les concepts jugés les moins importants ou les moins assurés par l'agent qui les possède.

Cela signifie entre autre que l'agent ne doit pas changer de concept à la légère, mais confronter le concept défaillant ou le problème soulevé par une impossibilité de représenter convenablement et adéquatement l'environnement, à l'ensemble des représentations et des concepts qu'il possède. Et cela non seulement pour préserver la consistance de son espace cognitif,

mais également pour préserver l'usage qu'il fait de ses représentations. Si les concepts permettent à l'agent de représenter son environnement afin de former des croyances, des connaissances, des désirs, des actions sur lui, l'agent doit tenter au maximum de préserver des croyances, connaissances, désirs et plans d'action, afin de rester cohérent avec lui même et son histoire. Sinon l'agent sera jugé inconséquent et inconsistant à changer d'avis et de représentation du monde à tout bout de champs.

De ce principe de préférence dans le changement découle le fait que l'agent doit préférer réviser le moins possible ses concepts i.e. le faire dans la mesure du nécessaire.

6.2.6 Principe de minimalisation dans le changement

Le *principe de minimalisation dans le changement* (PMC) dit qu'en plus du fait que l'agent doit préférer réviser les concepts les moins assurés selon lui, il doit tenter de réviser le concept uniquement dans la mesure où celui-ci peut rester consistant avec les autres représentations de l'agent.

(PMC) Le changement – dans l'intension ou l'extension – dans la révision du concept ne doit pas excéder le nécessaire.

Et ce dans l'hypothèse que l'agent est rationnel c'est-à-dire qu'il entend préserver au maximum son système de représentation et de compréhension du monde. Un agent qui changerait plus que de mesure un concept serait jugé comme n'apportant pas d'importance à ce concept. Or si le concept est une entité mentale qui représente l'environnement pour l'agent qui la possède, si un agent ne tient pas compte d'un concept ou le juge sans importance au point d'accepter de le passer par dessus bord à la moindre occasion, ou bien cela signifie que ce concept n'est pas jugé être un bon concept par l'agent et alors celui-ci aurait déjà dû le réviser auparavant, ou bien cela signifie que l'agent n'accorde pas d'importance à sa manière de représenter son environnement, et donc ses actions, désirs, connaissances et croyances passeront pour franchement douteuses.

Un agent rationnel doit accepter de changer de concepts lorsqu'il le faut, mais il est supposé ne pas aimer faire cela, et ne changer de concept qu'en dernier recours et avec parcimonie. Puisqu'en changeant de concept il change – même si ce n'est qu'un peu – de représentation de son environnement, i.e. sa vision du monde, et finalement il change de monde, ce qui n'est pas une mince affaire.

6.2.7 Principe de correspondance catégorielle

Le *principe de correspondance catégorielle* (*Principle of Categorical Matching* (PCM)), développé pour la révision des croyances, est formulé de la sorte par Rott, p. 44 :

(PCM) The representation of a belief state after a belief change has taken place should be of the same format as the representation of the belief state before the change ⁴.

Ce principe exprime le fait qu'une croyance doit avoir le même format avant et après le changement. Ce principe est sans doute trop fort, à moins d'une certaine libéralité sur le format. En fait il suffit que la nouvelle croyance – la croyance révisée – puisse être substituée pour l'ensemble des occurrences pertinentes de la croyance qu'elle remplace excepté pour les contextes qui ont conduit à la révision. Donc au lieu d'une correspondance exacte de catégorie – catégorie est ici à comprendre en tant qu'elle s'applique à un langage de la pensée – une substitution de l'ancienne croyance par la nouvelle est suffisant.

Dans le cadre inférentiel de la théorie proposée ici, la correspondance se comprend en terme de type. Ce principe vaut également pour les concepts. Un concept révisé doit pouvoir rendre compte des occurrences qui restent pertinentes du concept qu'il remplace. C'est-à-dire que si le concept α est révisé en le concept α^* , alors α^* doit pouvoir rendre compte de l'ensemble des usages et des applications de α moins les occurrences auxquelles α ne s'applique plus du fait de ξ dans l'environnement.

Ce principe est évident et raisonne comme une lapalissade. Si le concept révisé ne permet pas de retrouver les usages et les applications du concept

⁴“(PCM) La représentation d'un état de croyance après qu'un changement de croyance a eut lieu devrait être du même format que la représentation de l'état de croyance avant le changement.”

qu'il remplace alors ce concept apparaîtra comme un nouveau concept et non pas comme remplaçant le concept en question. La question se pose alors de savoir dans quelle proportion un concept est révisé et à partir de quel niveau de changement le concept révisé n'apparaît pas comme un nouveau concept. Il est possible de répondre au moins partiellement pour le moment à cette question. Dans la théorie inférentielle présentée ici le concept est défini à la fois par ses conditions d'usage – son intension – et par son application – son extension. Le concept révisé doit remplacer un concept, ce qui signifie que le concept révisé doit remplacer le nouveau concept à la fois intensionnellement – en permettant les mêmes inférences et relations avec les autres concepts et les justifications du concepts et le caractériser les mêmes occurrences que l'ancien concept et expliquer pourquoi il ne s'applique plus à certains objets et ne permet plus certaines inférences.

Admettons qu'un nouveau concept β – c'est-à-dire un concept qui n'apparaissait pas dans l'espace de l'agent auparavant – ait exactement les mêmes conditions d'usages et d'application que le concept α possédé par le l'agent.

$$\begin{array}{l|l} 1 & \lceil \Gamma \vdash \alpha \rceil \\ 2 & \lceil \Gamma \vdash \beta \rceil \\ \hline 3 & \lceil \alpha \doteq \beta \rceil \end{array}$$

c'est-à-dire que les justifications d'usage des concepts α et β sont les mêmes, alors il n'est pas possible de différencier ces concepts sur la base de leur intension. Et si

$$\begin{array}{l|l} 1 & \epsilon \alpha \doteq \epsilon \beta \\ \hline 2 & \lceil \alpha \doteq \beta \rceil \end{array}$$

c'est-à-dire que l'extension des concepts α et β sont les mêmes, alors il n'est pas possible des les différencier sur la base de leurs applications, et donc les concepts α et β sont identiques.

De deux choses l'une, ou bien ces deux concepts sont strictement identiques et alors par application du rasoir d'Ockham et par économie l'un des deux sera abandonné, ou bien alors l'un des concepts voit son intension et son extension faire partie de l'autre concept, mais ce dernier

a des conditions d'usages ou d'applications que ne possède pas le premier. Autrement dit l'un des concept est contenu dans l'autre qui est plus vaste. Alors le concepts subsumé par l'autre sera plus précis et plus spécifique que l'autre. Cela nous conduit à une vision hiérarchique du concept, exactement comme le concept CHIEN est contenu dans la concept ANIMAL. Mais alors ces deux concepts différents quant à leurs intensions – l'un a des conditions d'usages moins spécifiques que l'autres – et sur leurs extensions – le domaine d'application de l'un est plus restreint que l'autre, et donc ces deux concepts sont différents donc distincts.

Le cas qui nous intéresse, celui de la révision et du changement du concept, nous avons un concept α justifié par Γ et donc l'extension est $\epsilon\alpha$ dans l'environnement \mathcal{E} , et lorsque survient l'élément ξ dans \mathcal{E} le concept α ne s'applique plus pour une raison ou une autre. Or le concept α^* qui consiste en le concept α amendé de $\ulcorner \xi \urcorner$ – le concept qui décrit l'élément ξ – s'applique partout où s'appliquait α excepté pour les situations où apparaît ξ . Par exemple, le concept MAMMIFÈRE s'appliquait à l'ensemble des animaux qui allaitent le petits et qui ne pondent pas des œufs jusqu'en 1798 et plus ensuite du fait de l'apparition de l'ornithorynque dans l'environnement considéré. Mais si le concept MAMMIFÈRE tel qu'il apparaît jusqu'en 1798 est révisé relativement à la caractéristique de ne pas pondre des œufs tout en gardant celle d'allaiter ses petits, alors il permet de décrire l'ensemble des animaux qu'il décrivait jusque là et permet de décrire ce nouvel élément qu'est l'ornithorynque. Cet affaiblissement des conditions d'applications et d'usage ne modifie pas de manière conséquente la classification antérieure – e.g. les êtres vivants décrits avant 1798 à l'aide du concept OVIPARE ne seront pas décrits par le concepts MAMMIFÈRE, excepté pour l'ornithorynque mais qui n'apparaissait pas dans l'extension antérieure du concept OVIPARE, et aucun des membres de MAMMIFÈRE n'est perdu dans la modification du concept – et permet de rendre compte de ce nouvel élément. Dans ce cas le concept MAMMIFÈRE sera amendé et modifié en ce qui concerne son intension :

$$[\ulcorner \Gamma \urcorner \dot{\vdash} \alpha \urcorner, \xi \dot{\vdash} \neg \alpha] \rightarrow [\ulcorner \Gamma \wedge \varphi \urcorner \dot{\vdash} \alpha \dot{\vdash} \xi \urcorner] \quad (6.14)$$

(1) si ξ n'apparaît pas dans α et que par ailleurs ξ est contradictoire avec l'une des conditions d'applications de α , alors ou bien ξ ne doit pas être compris dans l'extension de α et doit être décrit par un autre concept et si aucun autre concept ne décrirait mieux ξ que α alors il faut créer de toute pièce un concept s'appliquant spécifiquement à ξ , ou bien, (2) si l'amendement de α par la condition φ permet de décrire ξ dans α sans modifier par ailleurs l'application de α ni d'aucun autre concept possédé par l'agent cognitif, alors il faut réviser α en l'amendant de φ . Dans l'exemple de l'ornithorynque, φ est la condition de ne pas pondre d'œufs. Si cette révision marche, alors il apparaîtra que la condition φ n'était pas essentielle – et donc a fortiori non suffisante pour justifier α .

Le mouvement peut être inverse, au lieu de modifier les conditions d'usages, i.e. intension du concept, il se peut que ses conditions d'applications soient changées. C'est le cas par exemple pour la baleine qui est retirée du groupe des poissons pour passer dans celui des mammifères. Dans ce cas, les baleines seront retirées du groupe des poissons uniquement par affinement de leur description dans le monde. Le concept MAMMIFÈRE peut avant et après la modification de son extension garder les mêmes conditions d'usage, mais l'affinement de la discrimination des objets décrits dans l'environnement – les baleines ont été observées de manière plus fine et il a été découvert qu'elle ne pondaient pas des œufs et qu'elles allaitaient leurs petits – et il s'est avéré qu'alors qu'il était pensé jusque là que le concept POISSON permettait de les décrire, en fait ces êtres vivants instancieraient les caractéristiques propres au concept MAMMIFÈRE.

Dans ce cas il est vrai qu'il ne s'agit pas à proprement parler de révision du concept, mais simplement d'un meilleur usage, d'une meilleure application de celui-ci.

En somme donc le changement des justifications correspond à un changement dans la trame conceptuelle, alors que le changement d'extension correspond à un changement dans le monde, du moins de la manière d'en rendre compte.

Chapitre 7

Concept et croyance

L'objet de ce chapitre est de comparer la théorie du changement de concept proposé ici avec la théorie AGM [Alchourrón et al., 1985] du changement de croyance.

La théorie AGM propose une formalisation de la révision dans le cadre de la théorie de la cohérence. Une croyance est révisable si elle se conforme à des critères de rationalité, en particulier respecter les principes de consistance et de fermeture logique. Trois opérations permettent de réviser une théorie, ce que AGM appelle l'*expansion*, la *contraction* et la *révision*.

7.1 Expansion

L'*expansion* est une opération qui consiste à ajouter un élément à un ensemble ne le comprenant pas, sans que cette adjonction modifie la structure ou les fonctions des éléments préexistants, ni leurs relations. En ce sens, le terme d'expansion est synonyme de prolongement ou de continuation.

7.1.1 Expansion de croyance

Dans la théorie AGM, l'expansion de croyances est définie à partir du moment où un état de croyance \mathcal{B} décrit par un ensemble de formule \mathcal{K} , et où A est une formule consistante avec \mathcal{K} qui est ajoutée à l'état de croyance \mathcal{B} , il en résulte un nouvel état de croyance \mathcal{K}_A^+ donné par la clôture logique de l'ensemble des formules $\{\mathcal{K} \wedge A\}$, soit, par l'intersection des modèles $\mathcal{M} \models \mathcal{K}$ avec l'ensemble des modèles $\mathcal{M} \models A$.

7.1.2 Expansion de concept

Il faut distinguer l'expansion du concept au niveau de son application de son expansion au niveau de son usage.

7.1.2.1 Expansion de l'application du concept

L'expansion de l'application du concept correspond au fait qu'un nouvel élément subsumé par le concept apparaît dans l'environnement dans lequel est utilisé le concept.

Si le concept α s'applique à l'ensemble d'élément $\{a_\alpha\}$ dans l'environnement \mathcal{E} au moment t_1 , et si a_n est un élément de \mathcal{E} à t_2 tel que $a_n \notin \{a_\alpha; \mathcal{E}; t_1\}$ mais $a_n \in \{a_\alpha; \mathcal{E}; t_2\}$, c'est-à-dire que ce nouvel élément est subsumé par α dans l'environnement dans lequel cet élément apparaît, alors α devient le concept α^* dont l'extension est l'union de $\{a_\alpha\} \cup \{a_n\}$.

Exemple : l'expansion de l'extension de concept est relativement triviale, il correspond simplement à la découverte par l'agent qu'un élément tombe sous l'extension d'un concept. Mathieu voit pour la première fois un Terre-Neuve, on lui apprend que c'est un chien ou bien il le classe de lui-même dans le concept CHIEN il semble en avoir les caractéristiques, alors Mathieu élargit le champ d'application de son concept CHIEN à Chinook, le Terre-Neuve qui lui fait face. Cette méthode d'élargissement de l'extension du concept est celle décrite par les théories extensionnelles du concept telle que la théorie des prototypes, des exemplaires ou la théorie empirique du concept.

7.1.2.2 Expansion de l'usage du concept

L'expansion de l'usage du concept correspond non pas à l'élargissement de son application mais de son sens, c'est-à-dire aux inférences qui lui sont accessibles. En somme, au lieu de modifier l'environnement en lui ajoutant un nouvel élément, c'est le concept lui-même qui est modifié en ajoutant une propriété à sa fonction caractéristique. Cela change son intension et par conséquent son usage.

Imaginons qu'au concept α corresponde la fonction caractéristique $\lambda\Phi$. Soit la propriété ψ telle qu'elle n'appartient pas à $\lambda\Phi$ mais soit consistante avec Φ , ajoutée à α . Le concept α devient alors le concept α^* dont la fonction caractéristique est $\lambda\{\Phi \wedge \psi\}$.

L'extension du concept change en conséquence du fait que les éléments subsumés par α doivent non seulement satisfaire les caractères de Φ mais également ceux de ψ . Seront donc rejetés les éléments subsumés par α qui ne satisfont pas ψ et ajouté à l'extension de α ceux décrits par ψ qui satisfont Φ . Cependant comme ψ doit être consistant avec Φ , et qu'il s'agit d'une expansion, le nouveau caractère n'est pas censé rejeter des éléments précédemment subsumés par α .

L'usage de α est modifié car s'ajoutent maintenant aux inférences qui lui sont accessibles l'ensemble de celles permises par ψ . Comme il s'agit d'une expansion de concept, l'arrivée de ψ ne doit pas faire que des inférences permises par α avant l'adjonction de ψ ne soit plus permis après cet ajout, sinon il ne s'agit pas d'une expansion mais d'un remplacement d'un concept par un autre.

Exemple : si le concept MARIAGE correspond à l'union entre un homme âgé de plus de dix-huit ans et d'une femme âgée de plus de seize ans [article 144 du code civil français], et qu'il est maintenant étendu à l'union entre deux personnes majeures sans distinction ou spécification de sexe, alors les inférences accessibles au concept MARIAGE sont modifiées. Il est possible par exemple de parler de mariage homosexuel. Cependant certaines inférences accessibles par ce concept avant son expansion, sont toujours accessibles, l'union entre une femme et un homme devenant un cas d'union entre deux personnes. La question du domaine pertinent d'inférences qui doivent être accessible à un concept pour le caractériser, se pose. En effet le concept ÊTRE PARENT est inférentiellement accessible à partir du concept MARIAGE lorsque ce dernier est compris comme étant l'union entre personne de sexes différents mais l'est de manière différente lorsqu'il s'agit simplement de l'union entre deux personnes.

Il faut noter que l'expansion de l'usage du concept passe en fait par l'affaiblissement de sa fonction caractéristique, la propriété ajoutée ne devant pas apporter une nouvelle spécification.

L'expansion du concept correspond à ce que [Putnam \[1973\]](#) appelle le « simple changement » (*'mere change'*).

7.1.2.3 Pourquoi parler de changement de concept plutôt que de remplacement ?

Par principe d'économie, si l'ensemble des concepts, de leurs applications et de leurs relations ne sont pas changés dans une conception, c'est-à-dire une trame conceptuelle, ne sont pas modifiés par l'ajout d'un nouvel élément ou d'une nouvelle propriété, alors on parle d'évolution ou de changement de concept plutôt que de remplacement de concept, qui affecterait la conception dans son ensemble.

Soit α et β deux concepts tels qu'à α corresponde la fonction caractéristique $\lambda\Phi$ et qu'à β corresponde la fonction caractéristique $\lambda\{\Phi \wedge \psi\}$, de sorte que ψ n'ajoute rien de spécifique à Φ et que $\lambda\Phi \in \{\Phi \wedge \psi\}$. Soit \mathcal{C} une conception dans laquelle paraissent α et β . Pour toute situation s , l'élément a est décrit également par α et β , et que β décrit l'ensemble des éléments décrits par α mais non l'inverse. Alors α n'a pas de condition d'application que β ne vérifie pas et ne peut vérifier les conditions d'application de β que si ψ lui est ajouté. Par principe d'économie, comme ce concept n'est jamais utilisé sans que soit simultanément utilisé β , α doit être abandonné au profit de β . Mais comme β est construit à partir de α auquel ψ est ajouté à sans fonction caractéristique sans qu'un autre concept de \mathcal{C} ne soit affecté, toujours par principe d'économie, mieux vaudrait parler d'évolution de α que de remplacement de α par β .

Contrairement à l'expansion de croyance, l'expansion de concept ne peut se comprendre en termes d'ajout de proposition à une théorie comme ensemble d'énoncés. L'expansion de concept ne correspond pas à la simple intersection de modèles. Il faut rendre compte des deux dimensions du concept. L'expansion d'extension correspond à une extension de modèle, mais l'expansion d'usage implique de prendre en considération les inférences permises par l'ajout du nouveau caractère. L'expansion d'application correspond à une *restriction* de la fonction caractéristique, qui devient moins spécifique. Si la conception n'est pas affectée par l'expansion du concept, on parle d'évolution ou de changement plutôt que de remplacement d'un concept par un autre.

7.2 Contraction

La *contraction* est une opération qui consiste à retirer un élément d'un ensemble le comprenant, sans que cette suppression modifie la structure ou les fonctions des éléments préexistants, ni leurs relations. En ce sens, le terme contraction est synonyme d'évolution ou de continuation.

7.2.1 Contraction de croyance

Dans la théorie AGM, la contraction de croyance est définie à partir du moment où à l'état de croyance \mathcal{B} décrit par un ensemble de formules \mathcal{K} , et où A est une formule non tautologiquement vraie qui est ôtée de \mathcal{K} . Il en résulte un nouvel état de croyance \mathcal{K}_A^- , logiquement clos, obtenu à partir d'une sélection de modèles de \mathcal{K} .

Cette sélection de modèles peut être *totale*, *partielle* ou *maximale*.

Contraction totale (*full meet contraction*) : est sélectionné l'ensemble des modèles de \mathcal{K} auquel A a été retiré, c'est-à-dire qu'est choisi l'ensemble des formules de $\mathcal{K} \perp A$.

Contraction partielle (*partial meet contraction*) : n'est retenu que le sous-ensemble strict des modèles de $\mathcal{K} \perp A$.

Contraction maximale (*maxichoice contraction*) : n'est retenu qu'un singleton vérifiant $\mathcal{K} \perp A$.

7.2.2 Contraction de concept

Tout comme pour l'expansion, il faut distinguer la contraction au niveau de l'application du concept de la contraction au niveau de l'usage du concept. La question se pose également de savoir s'il est effectivement possible de contracter un concept. Pour répondre à cette question il faut distinguer la contraction de concept, que ce soit de son extension ou de son application, et la contraction de conception, c'est-à-dire le retrait pur et simple d'un concept du répertoire d'un agent. Ce dernier cas s'apparente en fait à la révision du concept, et non pas à sa contraction qui n'est que la contraposée de l'expansion exposée précédemment.

7.2.2.1 Contraction d'application de concept

La contraction d'application du concept correspond à la suppression d'un élément de son extension, c'est-à-dire à une modification de l'environnement dans lequel le concept s'applique.

Soit le concept α dont le domaine d'application est l'ensemble d'éléments $\{a_\alpha\}$ et que l'élément a_n inclu dans $\{a_\alpha\}$ vient à disparaître. Il en résulte que le concept α devient le concept α^* qui s'applique à l'ensemble d'éléments $\{\{a_\alpha\} \wedge \neg a_n\}$. La contraction correspond donc au modèle vérifiant α auquel le modèle vérifiant a_n a été soustrait.

1. La contraction totale correspond à l'ensemble des modèles vérifiant α à l'exception de ceux qui vérifient a_n ,
2. La contraction partielle correspond au sous-ensemble strict des modèles vérifiant α à l'exception de ceux qui vérifient a_n ,
3. la contraction maximale correspond à un sous-ensemble vérifiant α mais pas a_n .

Exemple : Si un agent cognitif pense que les baleines sont des poissons, c'est-à-dire que du concept POISSON il peut en inférer le concept BALEINE, et qu'ensuite il pense que les baleines sont des mammifères, alors l'usage de son concept BALEINE change. Évidemment l'extension des concepts POISSON et MAMMIFÈRE changent également en conséquence, mais sans affecter l'ensemble de la conception de l'agent.

La contraction totale d'extension du nouveau concept POISSON de l'agent, correspond à l'extension précédente du concept, à laquelle les baleines ont été ôtées.

La contraction partielle d'extension du concept POISSON inclut les truites et les saumons.

Le choix maximal de contraction d'extension du concept POISSON subsume les truites fario et plus exactement celle qui frétille au bout de ma ligne.

La contraction extensionnelle du concept est triviale, elle correspond à la contraction des modèles exactement comme pour la contraction des croyances.

7.2.2.2 Contraction d'usage du concept

La contraction de l'usage du concept est moins évidente.

Soit le concept α et $\{\beta_\alpha\}$ l'ensemble des inférences accessibles par ce concept. Imaginons β_1 inclu dans $\{\beta_\alpha\}$ en \mathcal{E} à t_1 qui n'y appartient plus en t_2 . Le concept α devient alors le concept α^* en \mathcal{E}, t_2 dont le domaine d'application est $\{\{\beta_\alpha\} - \beta_1\}$.

Par exemple si l'agent pensait que les baleines étaient des poissons et qu'il pense maintenant que ce sont des mammifères, la propriété « être un poisson » a disparu de la fonction caractéristique.

La contraction totale correspond à l'ensemble des inférences permises par le concept BALEINE moins les inférences permises par le concept POISSON.

La contraction partielle d'usage du concept comprend les inférences permises par le concept BALEINE.

La contraction maximale correspond à une inférence permise par le concept BALEINE, par exemple : être l'animal qui a gobé Jonas.

Il faut faire attention à ne pas considérer la contraction d'usage du concept avec une révision de celui-ci. Dans le cas de la contraction, un caractère est éliminé, par exemple la propriété d'être un poisson est ôtée des caractères du concept BALEINE, et donc les inférences accessibles depuis cette propriété également, mais ce caractère n'est pas encore remplacé par un autre, par exemple celui d'être un mammifère. Pour cela il faut *réviser* le concept.

7.3 Révision

La *révision* est une opération qui consiste à vérifier des éléments dans un ensemble et leurs relations afin de les rectifier pour, au besoin les amender et les adapter aux circonstances, de l'expérience et de l'évolution du contexte.

7.3.1 Révision de croyance

La révision de croyances est définie par Levi [1977] en terme de contraction et d'expansion. Si \mathcal{B} est un état de croyance décrit par l'ensemble de formules \mathcal{K} , et si A est une formule inconsistante avec \mathcal{K} qui est ajouté à cet ensemble, il en résulte un nouvel état de croyance \mathcal{K}_A^* donné par l'identité

de Levi :

$$\text{(Identité de Levi)} \quad \mathcal{K}_A^* = (\mathcal{K}_{\neg A}^-)^+ \quad (7.1)$$

La révision de la croyance assurée d'aboutir à un ensemble consistant du fait que toute contradiction au sein de l'ensemble des formules définissant \mathcal{K} est évitée par l'opération de contraction. La contraction totale est cependant trop restrictive, dans la mesure où la contraction de la formule $\neg A$ avec l'ensemble \mathcal{K} ne conserve que les formules A qui découlent logiquement de A , c'est-à-dire que l'identité de Levi conduit à la formule :

$$\mathcal{K}_A^* = A \quad (7.2)$$

Ce qui est trop radical comme révision. Quant à la révision par la contraction maximale ('*maxichoice contraction*'), elle est trop restreinte dans la mesure où elle se limite à un singleton de A , sans trop savoir comment le justifier.

Toute la question consiste donc à savoir ce qu'il faut réviser dans un ensemble de croyance pour que cette révision soit opérante. Si l'on révisé trop ou pas assez, les conséquences sont négatives. Tout consiste donc dans le choix des éléments à réviser dans la croyance.

7.3.2 Révision de concept

Il faut distinguer la *révision de concept* de la *révision de conception*. La conception est composée d'un ensemble de concepts inférentiellement reliés les uns aux autres. La conception correspond à la *sémantique des rôles conceptuels*, et à ce niveau-là, l'extension du concept compte moins que ses relations aux autres concepts.

7.3.2.1 Révision de l'application du concept

La révision de l'application du concept est relativement triviale. Elle correspond à la révision extensionnelle, c'est-à-dire à l'ajustement du modèle vérifiant le concept compte tenu des éléments qui apparaissent ou qui disparaissent dans l'environnement. Cette application se fait comme un ajustement du modèle de manière similaire à la révision de croyance.

L'extension du concept à laquelle est ajouté ou soustrait les éléments en question est examinée : si elle est consistante et ne contredit pas ou n'est pas

identique à l'extension d'un autre concept présent dans la conception, alors la nouvelle extension est attribuée au concept. Si la nouvelle extension est inconsistante alors il faut adapter l'intension du concept jusqu'à ce qu'elle devienne consistante et non contradictoire ou redondante avec aucun autre concept présent dans la conception. Dans le cas où la nouvelle extension fait que le concept est identique à un autre concept déjà présent dans la conception, alors ou bien les deux concepts fusionnent – les termes lexicaux qui leur sont associés deviendront synonymes – ou bien l'un est délaissé au profit de l'autre. Ces cas sont expliqués au chapitre (6).

Cette révision de l'extension du concept peut amener à changer la conception et les relations entre les concepts. Si un concept est remplacé par un autre ou bien s'il est perdu au profit d'un autre, il est évident que les relations inférentielles accessibles depuis ce concept sont affectées et changent.

7.3.2.2 Révision d'usage du concept

La révision d'usage du concept affecte directement la conception puis qu'elle se fait au niveau même des relations inférentielles accessibles à partir du concept. Si donc la conception est caractérisée par les relations entre les concepts, tout changement d'intension d'un concept a des répercussions au niveau de la conception.

Pour réviser l'intension d'un concept, on considère son intension présente à laquelle est ajoutée ou retirée la propriété qui apparaît ou disparaît. Si la nouvelle fonction caractéristique est consistante avec l'ancienne et si elle n'est ni contradictoire ni redondante avec la fonction caractéristique d'un concept présent dans la conception, alors cette nouvelle fonction caractéristique est assignée au concept. C'est par exemple le cas pour un glissement sémantique tel que pour le terme 'mariage' qui s'appliquait à la simple union entre un homme âgé de plus de dix-huit ans et d'une femme âgée de plus de seize ans, à une union entre deux personnes majeures.

Si la fonction caractéristique à laquelle est ajoutée ou retirée la propriété en question devient inconsistante, alors elle doit être ajustée afin de devenir consistante, non contradictoire et non redondante avec un autre concept. Si le concept avec cette nouvelle fonction caractéristique est consistant mais contradictoire ou redondant avec un autre concept présent dans la conception, alors il faut réviser la conception elle-même, de la manière qui est exposée dans le chapitre (6).

7.3.2.3 Changement de conception

La question du changement de conception, c'est-à-dire du remplacement d'une conception par une autre à la suite d'une révision entraînée par un changement ou une révision de concept. C'est sous cette forme que le changement de concept est étudié en épistémologie et en philosophie des sciences, notamment par des auteurs tels que [Kuhn](#) [1962, 1970, 1996], [Feyerabend](#) [1980], en particulier en soulevant la question de savoir comment il est possible de comparer deux conceptions différentes – ce qu'ils appellent des *paradigmes* – s'il s'agit bel et bien de deux manières différentes de représenter le monde. [Kuhn](#) et [Feyerabend](#) sont partisans de l'*incommensurabilité* des théories, c'est-à-dire que deux conceptions ne peuvent être comparées, puisqu'aucun terme clé de l'une ne peut être traduit ou défini dans l'autre. Ce qui conduit à dire que toute théorie est relative et qu'il y a un particularisme ou un atomisme des théories.

Appliqué au concept, l'incommensurabilité des conceptions signifierait que, si chaque agent cognitif possède son propre répertoire de concept, c'est-à-dire sa propre conception, alors il ne pourrait communiquer aucune information ou aucune pensée à aucun autre agent cognitif possédant une conception différente de la sienne. En somme, postuler le concept revient à postuler le solipsisme.

Cependant, les principes de *référence canonique postulée* (PRPC) et de *dérivation canonique postulée* (PDC) affirment que la publicité du concept se base sur l'assignation par un agent d'une référence ou d'une dérivation compatible à la sienne possédée par un autre agent, ce qui revient à dire non seulement que l'agent *attribue* des concepts à autrui – ce qui suppose une théorie mentale de la part de l'agent – attribution qui rejoint la notion d'*interprétation* intersubjective proposée par [Davidson](#) [1984, 2001]. Interpréter, selon [Davidson](#) consiste à attribuer des croyances, des désirs et des intentions à autrui. Cela ne peut être fait de manière efficace que si l'agent est capable de comprendre correctement le « langage » d'autrui, c'est-à-dire s'il est capable de traduire ce langage dans le sien propre. En terme de conception, cela signifie que l'agent est capable de relier le concept d'autrui à son ou ses propres concepts. L'agent postule donc que son vis-à-vis est rationnel si sa conception est compréhensible par celle de l'agent, et cette compréhension revient à une forme de traduction de la conception de l'autre dans celle de l'agent. De fait, si l'agent utilise la postulation ca-

nonique pour interpréter et comprendre autrui, cela signifie radicalement qu'aucune conception, aucun schème conceptuel accessible à l'agent n'est incompatible et incommensurable avec le sien.

Putnam développe lui aussi un argument contre l'incommensurabilité des conceptions. S'il était vrai que deux conceptions étaient incommensurables, cela signifierait qu'il existerait un langage dans lequel rien ne pourrait être traduit, il ne serait qu'un bruit et alors il n'y aurait absolument aucune raison de postuler que ce qui produirait ces bruits pense ou parle ou raisonne. Mais au lieu d'en conclure immédiatement l'impossibilité *a priori* d'un tel langage, Putnam pointe du doigt sur le fait que l'interprétation historique serait incohérente. Dire que les notions utilisées par Galilée sont incommensurables avec les nôtres et de dire comment, revient à parler un langage qui n'exprime rien et donc être totalement incohérent. Si nous pouvons dire quelque chose de Galilée et dire en quoi notre conception du monde à changée, c'est que nous pouvons exprimer les notions qu'il utilise avec les nôtres, et donc qu'il y a une compréhension possible entre les conceptions, même si ce ne sont pas les mêmes, c'est-à-dire même si les concepts et les relations entre eux sont différents d'une conception l'autre. Reste à savoir maintenant comment la révision ou le changement de conception s'opère.

Chapitre 8

Conclusions et perspectives

8.1 Conclusions

Le concept est cette entité mentale qui permet à un agent cognitif qui la possède de représenter son environnement et d'avoir des états mentaux d'ordres supérieurs comme des croyances, des jugements, des connaissances, des désirs ou des plans d'action par exemple. Dire que le concept est une représentation c'est dire que l'agent doit pouvoir d'une manière ou d'une autre noter la distance qu'il y a entre le concept et l'élément de l'environnement auquel il l'applique. En d'autres termes, le concept est une entité qui s'attribue, et cette attribution est susceptible de correction. Si je dis que l'objet en face de moi est un ours blanc alors que c'est un renne, je représente l'objet qui est en fait un renne par le concept OURS BLANC, alors que je devrais le faire avec le concept RENNE. Le concept est différent d'autres entités mentales utilisées dans la relation que j'ai avec l'ours blanc en question. Par exemple je vois un certain objet, j'en perçois les formes, les couleurs et les mouvements, et je peux interagir correctement avec cet objet tout en pensant que c'est un ours blanc alors qu'en fait il s'agit d'un renne. Suivre ces objets du regard par exemple requiert de *discriminer* cet élément du fond sur lequel il évolue suivant mon point de vue par exemple, et cette discrimination peut être correcte et appropriée alors même que l'*identification* de cet élément échoue. Je peux suivre un renne des yeux et penser qu'il s'agit d'un ours polaire. En un sens cette explication du concept l'affaiblit un peu, puis que le concept se retrouve en concurrence avec d'autres entités mentales non-conceptuelles, comme

les percepts par exemple. Mais il ne s'agit pas de constituer une théorie pour elle-même, mais de rendre compte d'un phénomène mental particulier. L'une des caractéristiques principale du concept selon la théorie proposée ici est de considérer que le concept est une *entité représentationnelle*, alors que les entités non-conceptuelles qui entrent en jeu dans le traitement de l'information sensorielle ne le sont pas, du moins dans le sens de possibilité de correction. Pour le dire rapidement, lorsque je perçois un renne et que je pense qu'il s'agit d'un ours polaire, ma perception de l'objet est appropriée ou non, au sens où je vois que c'est un renne uniquement si certaines conditions physiques – la luminescence, le fait d'avoir les yeux ouverts par exemple – sont respectées. De fait la relation entre le percept et l'objet est nomologique et causale, et tient plus de la pertinence des propriétés physiques traitées de l'objet que de la vérité ou de la fausseté de ce traitement. Le concept au contraire est la manière dont l'agent cognitif pense le monde, et si je pense que l'objet que je perçois est un ours polaire alors qu'il s'agit d'un renne, je perçois bien un objet mais je me trompe dans son identification. Dire que je me trompe implique que je puisse d'une manière où d'une autre corriger mon erreur. Cela suppose que je puisse avoir accès au contenu et à l'usage de ma représentation. L'accès au contenu ou à l'usage de la représentation reste au niveau de la représentation. Lorsque je me trompe ne prenant le renne pour un ours, je représente bien quelque chose et interagis bien avec cette représentation. C'est dans ce comportement et dans les relations entre les représentations et la perception de l'environnement que l'erreur prend sa valeur. Si je prends mon fusil et que je mets en joue le renne que je prends pour un ours blanc parce que je pense qu'il est dangereux, mon comportement peut être efficace – si je tire je tue un animal – mais pour une raison que n'est pas la bonne. Je tire parce que je crois l'animal agressif, alors que le renne ne l'est pas. Mon action est belle est bien une action, mais motivée par une raison qui n'est pas bonne.

C'est parce que le concept est une représentation, et donc qu'il est susceptible d'être mal appliqué ou mal utilisé qu'il peut être corrigé, et c'est parce qu'il peut être corrigé que le concept doit être susceptible de changer. D'où l'importance du changement conceptuel dans l'étude du concept. Mais comment un agent peut-il changer de représentation s'il ne peut sortir de la représentation ? Cet argument sceptique et relativiste dit quelque chose

comme « comment pouvez vous parler de changement au sein du concept si vous ne pouvez pas comparer l'ancien et le nouveau concept ? ». Mais cette incommensurabilité ne fait pas justice du fait que l'agent peut être conscient de ce changement et qu'il peut l'opérer lui-même volontairement. Je suis d'accord avec le fait que le changement dans le concept se fait bien au niveau de la représentation, mais le concept est une entité complexe pourvue au moins de deux types différents de contenus : le contenu privé ou *cognitif* et le contenu public ou *canonique*. Ce que souhaiterait peut-être un réaliste externe radical serait que l'agent puisse avoir accès au contenu réel ou *objectif* du concept, c'est-à-dire non pas à la représentation mais à la propriété de l'objet perçu. Si je vois un renne, je dois avoir la possibilité de savoir que c'est un renne en vertu même du fait que c'est un renne. Évidemment un tel accès serait très pratique, en particulier pour constituer une épistémologie fiable, le seul problème est que si l'agent a accès directement à la propriété en tant que représentation alors il n'y a pas de possibilité d'erreur, or il se trouve que les agents cognitifs se trompent parfois. Si le réalisme externe radical avait raison, alors il n'y aurait rien de telle que l'[illusion de la constance de taille](#) (cf. figure 2.1) par exemple. Or les agents cognitifs sont sujets aux illusions perceptuelles, et il faut bien en rendre compte, même si cela rend moins simple et moins belle la théorie.

La solution proposée au problème de l'évaluation et de la visibilité de l'erreur d'attribution ou d'usage du concept est de dire que l'agent a accès d'une part aux propriétés de l'objet – à travers un traitement non-conceptuel de l'information perceptuelle – qu'il a accès à sa propre représentation – l'attribution du concept α à l'objet o_1 – et aux raisons qu'il pense lui permettre d'attribuer ce concept à cet objet – l'application du concept α à l'objet o_1 en vertu de Γ – ainsi qu'aux raisons qu'il pense qu'un autre agent que lui invoquerait pour appliquer ce concept, d'autre part – la fameuse référence au contenu canonique.

Dit ainsi cela paraît bien abstrait, mais c'est bien ce qui se passe lorsque par exemple vous planifiez une expédition dans telle vallée du Svalbard et que vous voulez savoir quelle est la probabilité que vous avez de rencontrer des ours et combien, pour ajuster au mieux votre équipement. Vous interrogez les personnes qui vous semblent le mieux à même de vous renseigner, et vous vous rendez compte que chacun à sa version de l'histoire,

certaines vous disent qu'ils n'y ont jamais croisé un ours et d'autres qu'il y en a par dizaines. Qui croire et comment établir un critère d'évaluation entre ces témoignages? L'une des manières de faire, ce me semble, et de demander à chacune des personnes interrogées ce qu'elle pense que penserait quelqu'un d'autre qu'elle, par exemple ce qu'elle pense que quelqu'un qui voudrait mener une expédition dans cette région devrait prendre comme équipement. Assez étrangement, alors que chacun racontait *son* histoire et son expérience des ours dans ce coin, lorsqu'elle se met à penser ce que penserait un tiers, généralement toutes les personnes consultées s'accordent pour donner une même réponse, en faisant référence au *sens commun*, en répondant quelque chose comme « moi je n'en ai jamais rencontré, mais je dirais que quelqu'un qui doit aller là-bas devrait s'attendre à rencontrer un ours ou deux, et donc de prendre une carabine ». Même si quelqu'un vous dit qu'il en a croisé des dizaines vous dira que les chances d'en rencontrer plus de deux ou trois sont très faibles. Cette dissociation qu'est capable d'opérer l'agent entre son propre contenu d'un concept et le contenu en troisième personne, ou canonique du concept lui permet d'ajuster et de réviser son concept au besoin.

La révision du concept intervient dans un contexte de communication et d'échange d'information entre agents, mais également du fait que le concept à son utilité, en particulier pour constituer des états mentaux d'ordres supérieurs. La croyance selon laquelle les ours polaires sont dangereux repose non seulement sur le comportement effectif de certains animaux du monde qui sont des ours polaires, mais également sur l'évaluation ou la conception du danger par un agent. Il est probable que si vous parlez à quelqu'un qui se baigne sans aucune protection au milieu de requins et qui vous soutient que les ours polaires ne sont pas dangereux, vous allez considérer son jugement comme étant singulier et peut être vous dire que vous n'avez pas la même notion du danger que cette personne. L'inverse est vrai si quelqu'un vous dit qu'ils sont très dangereux tout comme le sont les hérissons. Le changement du concept s'opère grâce au fait que l'agent possède deux contenus au moins pour un même concept.

Pour résumer dans les grandes lignes, l'histoire va ainsi : il y a un environnement – le monde – et un agent cognitif qui évolue dans cet

environnement. L'environnement est meublé d'objets divers, des arbres, des chaises, des dahus et des thèses de philosophie. L'agent cognitif dispose d'une relation avec cet environnement par le biais de capteurs d'information, les sens. Ces sens détectent des propriétés qu'ont les objets de l'environnement, selon que certaines conditions sont remplies, par exemple il faut qu'il ait un objet porteur de la propriété détectée par le capteur dédié dans le champ du capteur pour que celui-ci se déclenche, il faut également que les conditions minimales pour le déclenchement soient respectées, par exemple que la luminance soit suffisante pour distinguer les couleurs lorsqu'il s'agit d'un capteur visuel. Parmi l'ensemble des informations captées et traitées par les capteurs de l'agent, certaines seulement seront considérées comme intéressantes pour l'interaction entre cet agent et cet environnement. Il s'agit des propriétés saillantes. La scène perçue traitée selon ces propriétés saillantes est simplifiée mais permet à l'agent d'évoluer dans l'environnement. Par exemple, dans la figure 2.1 les détails du fond sont perçus mais ne sont pas conservés lorsqu'il s'agit de comparer la taille des deux personnages. Ce traitement des informations perçues permet déjà une interaction très fine avec l'environnement. Des objets sont discriminés – même s'ils ne sont pas identifiés – de telle sorte qu'une action sur eux est possible. Par exemple il est possible d'éviter un obstacle en opérant un certain nombre de mouvements complexes, sans devoir identifier l'objet évité. Toutes ses opérations se passent au niveau non-conceptuel.

En parallèle au traitement non-conceptuel, l'agent *représente* son environnement. Il peut se baser sur les données non-conceptuelles de la perception ou bien représenter l'environnement indépendamment de ces données. L'agent peut isoler et représenter un élément de par les propriétés et les caractéristiques que possède cet élément. Il décrit ainsi l'élément. Il peut également assigner un contenu à une description, ou bien en mettant en relation cette description avec un élément du monde, ou bien en postulant que cette description décrirait quelque chose si certaines conditions étaient remplies. Il s'agit d'une part du contenu cognitif, et d'autre part de la dérivation cognitive. Le contenu cognitif est l'ensemble des informations que l'agent rapporte à une fonction caractéristique, la dérivation cognitive est l'ensemble des raisons qu'utilise l'agent pour faire une assignation.

Cependant l'agent cognitif n'est pas isolé. Il évolue dans un environnement peuplé l'objet mais également d'agents cognitifs différents et indépendants de lui. Du fait qu'une communication spécifique est possible entre les agents au moins d'une même espèce, un échange d'information entre agents est possible. Pour une part cette communication est conceptuelle. Se pose alors la question de l'adéquation entre les représentations, c'est-à-dire entre les concepts, de différents agents. Cette contrainte sur la représentation oblige l'agent à comparer le contenu et la dérivation qu'il attribue personnellement à un concept avec celles que pourrait faire un agent différent de lui avec lequel il pourrait communiquer. Ce sont les contenus et les dérivations canoniques, c'est-à-dire les contenus et dérivations qu'un agent pense en troisième personne. Rapporé au niveau de la communauté des agents, ces contenus et dérivations canoniques sont représentés par le sens commun. Ces relations entre les contenus et les dérivations impliquent une forme de holisme, au moins partiel, qui apparaît à travers les domaines des concepts, notamment dans les théories naïves. Les relations entre les concepts forment une trame conceptuelle qui correspond à une conception. Ces conceptions sont des groupements de concepts interconnectés. Suivant le type de connexions entre ces concepts, cela forme des états mentaux d'ordres supérieurs tels que les croyances, les jugements, les désirs, les plans d'actions, etc. Par exemple, dans l'image 2.1, la pensée, le jugement ou la croyance selon lequel les deux personnages sont de tailles différentes reposent dans un premier temps sur une mauvaise évaluation due à une interférence entre des propriétés saillantes au niveau de la perception. La correction de ce jugement se fait sur la base d'un examen des conditions et des relations entre les objets en fonction d'une information qui n'est pas directement donnée dans et par la perception mais qui est accessible au niveau conceptuel.

8.2 Perspectives

La théorie des concepts présentée est philosophique. Elle est plus normative que descriptive. Ce qu'il reste à faire est donc de voir dans quelle mesure ses prédictions – en particulier sur le rôle du sens commun et des théories naïves sur les concepts possédés par un agent cognitif – sont conformes avec l'activité cognitive d'un agent. Il s'agit donc de mener une investigation psychologique et expérimentale du concept sur la base proposée ici.

Un second axe de recherche est de poursuivre la formalisation logique de la théorie afin d'obtenir un système axiomatique pour comprendre la dynamique du concept comparable à celui dédié à la révision de croyance.

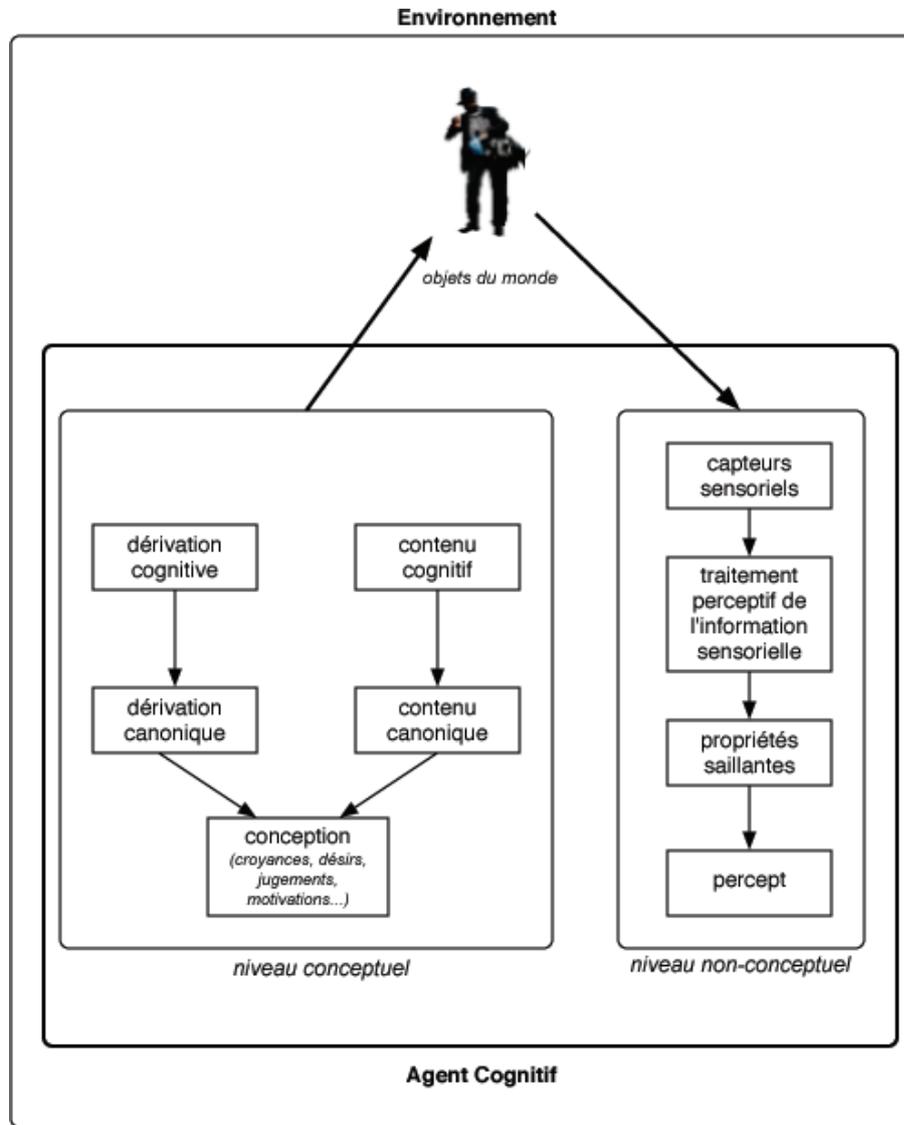


FIG. 8.1 – Schéma récapitulatif de la structure du niveau conceptuel et du niveau non-conceptuel

Quatrième partie

Annexes

Chapitre 9

Connaissance et incertitude.

Ce que le théorème de Ramsey nous apprend de notre connaissance.

[Texte d'une conférence donnée le 25 septembre 2001 à Aix-en-Provence]

Peut-on tout savoir ? Vaste question n'est-ce pas, qui même, posée ainsi, fait sourire. Non, la réponse est évidente et simple, non, nous ne pouvons pas tout savoir. Pourquoi ? Mais parce que nous ne sommes que de simples créatures finies, et que nous ne sommes pas omniscients, voilà pourquoi. Sans doute que Dieu ou le démon de Laplace peut tout connaître, peut-être, mais je n'en sais rien. Non, nous autres hommes ne pouvons pas tout connaître, et cela est si évidemment, si simple et si commun que tout le monde le sait. On nous reprocherait presque d'avoir posé une question aussi stupide.

Bien, mais comment savons nous que nous ne pouvons pas tout savoir ? Et bien, nous racontera probablement notre savant interlocuteur, parce que nous ne sommes pas dieu, ou nous rétorquera par une énigme du style « tu ne me crois pas qu'on ne peut pas tout savoir ? Et bien, pour t'en convaincre, essaie donc de me dire combien j'ai de cheveux sur le crâne, ou combien font 2345^2 , et si tu ne parviens pas à me répondre, alors je t'aurais prouvé que l'on ne peut pas tout savoir ! ». Il va s'en dire que cela ne prouve pas grand-chose, parce que je peux très bien employer une armada de coiffeuses qui vont lui épiler le crâne jusqu'à ce que je puisse les lui compter ses cheveux, ou bien prendre une calculatrice et lui répondre du tac au tac : 5499025,

pour que son beau raisonnement tombe comme un château de carte dans un courant d'air.

En effet, il ne s'agit pas de confondre entre la possibilité pratique de connaissance, et sa possibilité théorique. Evidemment que de tête et du tac au tac, j'aurais bien du mal à lui dire combien fait 2345^2 , surtout au vu de mes compétence mathématiques et de calcul mental, mais s'il m'est impossible pratiquement et sans l'aide d'une calculatrice de savoir cela, cela ne l'est pas théoriquement, la preuve et que je peux lui donner le résultat. Bien sûr, s'il me demandait de calculer chiffres immenses, j'aurais plus de mal, ma calculatrice ne serait plus assez puissante, mais théoriquement, cela ne serait pratiquement pas impossible.

Non, la réponse que l'on attendrait serait plutôt une réponse théorique. Peut-on tout savoir? Je ne vau pas savoir si je ne sais pas tout du fait de ma paresse, mais du fait même de la connaissance possible. Peut-on répondre à cette question?

C'est à cette question que j'aimerais répondre aujourd'hui. Mais il faut déjà tempérer mon propos. Non, ce soir je ne vais pas vous donner la réponse noir sur blanc à la question si toute connaissance théorique est possible. Désolé de vous décevoir. Non, mais je voudrais tout de même amorcé un début de raisonnement qui s'aventure sur la voie de cette réponse. En gros, il s'agit de savoir si la condition d'une connaissance absolument certaine est possible. Bien sûr, la manière dont je formule la question vous indique déjà mon point de vue sur la question. Mais ce n'est là qu'une opinion, la mienne en l'occurrence, et non pas une preuve. Essayons d'apporter une preuve un peu plus consistante.

Tout d'abord, examinons la question de la théorie de la connaissance.

Comme ce n'est pas principalement la question que je voudrais examiner aujourd'hui, je vais accepter la théorie classique, qui dit que X sait que p si et seulement si p est vrai, X croit que p et que X à des raisons de croire que p . Et je suivrais Ramsey pour dire que cela suppose que, pour qu'une croyance soit une connaissance, il faut qu'elle soit vraie, certaine et obtenue par un processus fiable. « Vrai » renvoie ici au fait considéré, c'est-à-dire ' p ', « certain » à la croyance que le sujet a de ce fait vrai, et « processus

fiable » à la vérification que le sujet peut avoir de sa connaissance certaine du fait vrai.

Cela implique trois choses donc, à commencer par une théorie de la vérité. Je ne développerais pas ici de théorie de la vérité, et je pense que pour le sujet qui nous intéresse ici, il est possible d'accepter n'importe laquelle des théories de la vérité, cela n'influe en rien sur le second problème qui est celui de la certitude de la croyance en le fait vrai examiné.

Nous pouvons considérer que nous avons des théories de la vérités qui sont satisfaisantes. Elles sont différentes et parfois contradictoires, mais cela ne nous importe pas trop pour le moment. On peut par exemple se contenter d'une théorie correspondantiste de la vérité et vivre très bien en paix.

Mais de savoir qu'un certain fait est vrai est-il suffisant pour savoir si notre connaissance en ce fait est fiable ou non? Ramsey propose ici l'usage d'un processus fiable et performant, c'est-à-dire qui marche à tout les coup et non pas au petit bonheur la chance pour savoir si le fait est bien vrai. Et il reste la question de la certitude. En effet, ce n'est pas parce qu'une connaissance porte sur un fait vrai, qu'elle est certaine. Prenons un exemple. Pierre croit que la Terre tourne autour du Soleil. Ce fait est vrai, i.e. il est vrai que la Terre tourne autour du Soleil, nous pouvons le vérifier par quelques expériences. Mais que le fait soit vrai, influe-t-il sur la certitude de la connaissance qu'à Pierre de ce fait? Non, assurément non. En effet, c'est peut-être Paul que le lui a dit, mais Paul raconte parfois des histoire qui ne sont pas vraies, il lui arrive de se tromper. La consultation de Paul n'est donc pas un processus fiable de connaissance. Pierre peut peut-être alors aller consulter une encyclopédie, mais même si cela paraît plus fiable ce ne l'est pas plus, l'encyclopédie peut-être ancienne et biaisée, ou être contrôlée par une secte révisionniste ou ne je sais quoi. Il faudra donc trouvé un critère de certitude qui soit donc plus fiable que cela. Y en a-t-il un?

L'un des critères qui semble le plus fiable dans une théorie de la preuve est celui de la réduction. Réduction à l'objet pour une connaissance empirique, aux axiomes pour une théorie. Le réductionnisme est-il possible au niveau de la certitude?

C'est sur cette question qu'il faut maintenant se pencher. Nous pourrions la formuler également ainsi : peut-on trouver un critère de certitude de la vérité d'une connaissance ?

Admettons, pour nous simplifier la tâche que nous ayons déjà une théorie de la vérité qui puisse nous permettre de répondre convenablement à la seconde partie de la question, c'est-à-dire, sur la question de la vérité. Peut-on être certain d'une connaissance.

Pour répondre à cette question, je vais proposer ici une interprétation du théorème de Ramsey, appliqué à la théorie de la connaissance. Il ne s'agit là encore que d'une ébauche de raisonnement, et il faut être très prudent. Les conclusions sur lesquelles débouchent ce raisonnement sont des plus surprenantes et des plus troublantes, mais elles sont, je crois, vraies. Avant même des les exposer, je répète la prudence avec laquelle nous devons les utiliser. J'apporterais ensuite des raisons à cette utilisation, mais cet appel à la prudence n'est pas une mise en garde contre le raisonnement que je vais présenter, mais contre toute éventuelle récupération inappropriée du théorème de Ramsey. Il faut à tout pris éviter la ramseyite comme il faut éviter la gödelite. Ce sont des maladies mortelles pour la raison. Mais je sais que vous êtes raisonnables, et que cet appel ne s'adresse pas à vous. Ceci dit, avançons.

Tout d'abord, le théorème de Ramsey .

Théorème A (version infinie) : soit T une classe infinie, et m et r des entiers positifs ; et $C_i (i = 1, 2, \dots, m)$ le nombre m de sous-classes de T mutuellement exclusives comprenant exactement le nombre r de combinaisons de membres de T , de telle sorte que chacune de ces r -combinaisons appartiennent à une et une sous-classe C_i . Alors, par l'axiome du choix, T contient une sous-classe infinie D telle que toutes les r -combinaisons de termes de D appartiennent toutes à la même C_i .

Qui s'écrit : $\omega \rightarrow (\omega)kr$

Et dans sa version finitiste :

Pour chaque entier positif fini k , r , et l , il existe un entier positif fini n qui vérifie la relation $n \rightarrow (l)kr$

Où (l) est un sous-ensemble de n , divisé en k sous-classes mutuellement exclusives de r membres chacune.

Nous pouvons trouver un élément x_1 et une sous-classe T_1 de T , telle que T_1 ne contienne pas x_1 , mais que x_1 et T_1 appartiennent tout deux à la même sous-classe C_1 de T . Et nous pouvons trouver dans T_1 un élément x_2 et une sous-classe T_2 qui ne contienne pas x_2 mais qui est comprise avec x_2 dans la sous-classe C_1 . Nous pouvons poursuivre cette opération autant de fois que nous le désirons, si bien que nous nous retrouverons à la fin avec deux séquences infinies de termes de T , l'une $T(x_1, x_2, \dots, x_n, \dots)$ des membres de T , et une autre, $T(T_1, T_2, \dots, T_n)$ des sous-classes de T , appartenant toutes deux à la même C_1 de T . Si on poursuit ce processus de division et de classification des éléments de T en autant de sous-classes correctement formées, nous rencontrons un paradoxe. En effet, nous examinerons à un certain moment la sous-classe T_{n-1} , qui ne contiendra pas l'élément x_n et qui contiendra la sous-classe infinie T_n , qui ne contiendra pas x_n pourtant appartiendra avec x_n tous deux à la même classe C_1 de T . Donc la classe C_1 de T contiendra une combinaison de terme dont l'un des membres n'est membre d'aucune des sous-classes de C_1 .

En remontant le processus à rebours, nous nous retrouvons avec un ensemble T , et un membre surnuméraire de T qui appartiendra pourtant à l'une des sous-classes de T , ou bien T contiendra une sous-classe qui aura le même nombre de membre que les autres sous-classes de T mais qui sera vide, alors que T ne l'est pas. Cela ne signifie pas que T soit mal constituée, mais simplement que sa division en sous-classe n'est et ne peut pas être exhaustive pour comprendre l'ensemble des combinaisons des termes de cet ensemble.

Ce paradoxe est étrange et fort troublant. Exprimé sous forme de graphe, il indique qu'il reste toujours un chemin possible non exploité entre les membres d'un ensemble, bien qu'ils soient tous connectés les uns avec les autres.

Appliqué à la théorie de la preuve, ce théorème nous indique qu'une procédure de vérification de connexion de membre d'un ensemble délaie toujours au moins l'examen d'une de ces connexion. Sorte de théorème d'in-

complétude de la théorie de la preuve. Le théorème de Ramsey est proche dans ses conclusions du second théorème d'incomplétude de Gödel, comme le remarquent [Paris and Harrington \[1977\]](#).

Le paradoxe que pointe le théorème de Ramsey, n'est pas un paradoxe ensembliste, parce qu'il distingue toujours nettement entre les ensembles et les éléments de ces ensembles, mais est plutôt un paradoxe d'incomplétude manifeste d'un processus de subsomption par catégorisation, c'est-à-dire un paradoxe touchant à la prédication même. Si l'on distingue les propriétés d'éléments et ces éléments, alors, en fin de compte, dans un système, même infinie de prédicats possibles, il restera toujours, potentiellement, un objet qui ne sera pas compris par ce prédicat mais qui pourra être visible. Cela va de soit si on se penche un peu sur l'affaire. Il va de soit que toute théorie de concepts est inachevée et ne peut être exhaustive, mais cela est tout de même troublant.

Revenons à notre problème de départ. Admettons que nous ayons une théorie de la vérité convenable et que nous acceptons. Celle-ci n'est pas suffisante pour former une théorie de la connaissance, puisqu'elle ne permet pas d'évaluer le degré de certitude de la croyance que nous avons de cette vérité. En évaluant la certitude d'un fait, nous devons passer à un niveau supérieur de la théorie, et ce niveau supérieur doit être compris dans une théorie plus globale qui comprend également la théorie sur laquelle porte cette certitude sans qu'elle contienne cette certitude. En fin de compte, nous nous retrouvons donc avec une connaissance qui n'est pas évaluable en terme de certitude et qui pourtant porte sur un fait vérifiable du monde.

Poussé à son terme, ce raisonnement nous indique que nous ne pouvons trouver les fondements de la certitude de la connaissance sur les mêmes bases que ceux de la vérité. Et que si nous admettons la connaissance comme étant une théorie unie alliant la vérité et les raisons de la croyance d'un sujet en cette vérité, alors cette théorie ne peut exhaustivement déterminer qu'elle est la certitude ou celle de la vérité. En admettant que nous ayons une théorie convenable de la vérité, cela implique que nous ne pouvons pas en avoir une, sur les mêmes bases, de la certitude.

Compris ainsi, le théorème de Ramsey indique l'incomplétude de la certitude de la théorie de la connaissance. Ce qu'il faut en conclure, c'est que nous ne pouvons n'être qu'incertain de notre certitude, ou certain de

notre certitude.

Je ne crois pas que cette interprétation du théorème de Ramsey soit en contradiction avec sa théorie de la connaissance, bien au contraire. Ramsey affirme qu'une croyance ne peut être tenue pour une connaissance uniquement si elle porte sur un fait vrai, et que nous ne soyons certains sur la base d'un processus fiable. Mais il met en doute la possibilité d'avoir un tel processus, car il reste toujours la possibilité que nous nous trompions, et comment pouvons nous le savoir? Il réfère alors à l'exemple du journal de Russell, ce n'est pas parce que nous croyons qu'une information est vraie en utilisant un processus qui nous paraît fiable, que cette croyance est une connaissance puisque nous pouvons toujours être sujet à l'erreur. C'est en fait comme de trouver quelque chose par l'internet, l'information peut être vraie et nous pouvons le croire, mais nous n'avons pas moins d'en être absolument certain, tout juste pouvons nous croiser les sources estimer de la fiabilité du processus suivant des probabilités ou des statistiques, mais cela ne garantit en rien de la certitude.

Dans un court texte en forme de note, intitulé « Theory of Knowledge » laisse entendre un tel scepticisme quant à l'évaluation du processus de certitude de la connaissance. On peut très bien admettre une théorie réaliste de la connaissance, cependant nous ne pouvons pas savoir si réellement la relation entre notre croyance en le fait et le fait opère vraiment. Donc nous considérerons cette connaissance comme vraie et certaine uniquement sur la foi que nous accorderons en la relation entre notre croyance et le fait. Mais il en va de même pour une connaissance internaliste. Il en conclu que la connaissance ne peut être garantie.

Il faut rappeler que cela ne n'affecte pas la vérité de la connaissance, mais uniquement l'évaluation de la certitude de son processus vérification. Tout comme le théorème de Gödel n'affecte en rien l'arithmétique, mais simplement la prouvabilité dans son propre système.

Il en va de même pour le théorème de Ramsey, ce n'est pas parce que je n'ai pas de preuve totale des processus que j'utilise que je dois passer toutes les lois de la logique par-dessus bord. Cela serait absurde.

Le théorème de Ramsey justifie plutôt pourquoi la machine universelle de Turing s'emballe parfois.

La conclusion du théorème de Ramsey est donc des plus troublante, elle nous indique simplement que nous ne pouvons prouver l'incertitude ou la certitude d'une connaissance. Mais il ne faudrait pas profiter de ce résultat pour tout délaissé en faveur d'un relativisme dogmatique. Dire que nous ne pouvons prouver la certitude d'une proposition vraie, que chacun puisse choisir ce qui est vrai ou non. Non, le relativisme de Ramsey est similaire au relativisme de Gödel, c'est-à-dire quasiment nul. L'incomplétude de la connaissance est relative non pas à chaque sujet humain, mais à tout Homme, tout spécimen de l'espèce. Nous sommes peut-être des cerveaux dans un cuve, peut-être, mais Ramsey nous indique que nous ne pourrons jamais le savoir avec certitude. Un point c'est tout.

Il nous reste alors à développer les outils de la connaissance qui sont les plus fiables et les plus utiles pour nous. Et ce résultat est un encouragement : s'il indique que la modélisation totale et panlogique est logiquement impossible, si les fictions fantasmées d'une intelligence humaine ou artificielle suppléant l'humain est autrement improbable et du moins ne sortira pas de l'Homme, le théorème de Ramsey ne nous dit rien d'autre que l'internet n'est qu'un moyen de communication comme un autre, certes performant, mais qui ne manipule que notre paresse et incompetence, mais n'est pas un super système qui nous échappe.

Mais cela n'est que bon sens et cela n'est que l'ébauche d'un raisonnement qu'il faudra approfondir.

Chapitre 10

Tracking Agency

[Article écrit en c-auteur avec Ángeles Eraña et Dario Taraborelli, soumis à la date de la rédaction.]

10.0.1 Abstract

The aim of this paper is to propose an extension of the *object file* notion to the study of nonconceptual individuation of *agents*. Robust evidence in both developmental and empirical psychology supports the hypothesis that dedicated perceptual mechanisms mediate sensitivity to *objecthood* as opposed to *animacy*. Object individuation and perceptual animacy have been largely studied in recent literature, whereas little is known about mechanisms mediating individuation and tracking of perceptual entities endowed with agency (or *proto-agents*). By introducing a notion of an *agent file*, we aim to provide a theoretical framework for more constrained empirical investigations into the ability to perceptually track agency (prior to explicit identification/categorization), as well as asymmetries with ordinary object tracking. We insist, in particular, on the need to understand mechanisms underlying the ability to represent agents' *persistence* beyond mere agency cue detection.

10.1 Individuation : Creating and Maintaining Reference

A large number of cognitive skills rely on the perceptual ability to single out individuals. In order to interact with, ascribe properties to, or reason about particular entities, we need to be able to *pick out* individuals, *establish* a referential link with them and *maintain* it over time. Many authors have pointed out that such ability must be grounded in the deployment of some nonconceptual skills : for picking out and maintaining reference to an individual, simply detecting some of its perceptual properties is not sufficient. We need to postulate a referential mechanism that provides a direct (nonconceptual and unmediated) link to this individual in order to be able to ascribe perceptual properties to it or to make perceptual judgments about it.

10.1.1 Reference to Objects *vs.* Reference to Agents in Infants

There is large evidence in the developmental literature indicating that, from early on, infants are able to deal with two distinct classes of individuals.

On the one hand, they are able to interact with and reason about *objects*, i.e. individuals that behave according to physical constraints. Understanding the observable behavior of objects requires — at least *prima facie* — a grasp of some of the principles underlying physical phenomena.

On the other hand, infants can interact with and reason about *agents*, i.e. entities endowed with intentionality, whose observable behavior cannot be reduced to those constraints governing objecthood. Being able to detect agency requires being sensitive to some specific observable cues of intentionality, such as purposefulness.

A prevailing working hypothesis in developmental studies claims that, when dealing with entities belonging to each of these domains (objects *vs.* agents), different individuation mechanisms are at work. Paul Bloom [Bloom, 2004] asserts that infants are *commonsense dualists* : ‘who have two ways of looking at the world : in terms of bodies and in terms of souls’ (p.191). Kuhlmeier et al. [2004a] have designed and carried out a number of experiments to support this idea : infants display precocious abilities to distinguish inanimate objects (entities that behave according only to physi-

cal constraints) from animate entities (e.g. humans). These abilities suggest that infants' early understanding of animate entities does not rely on the typical individuation principles and constraints on which the individuation of objects depends (p.7).

Most studies have thus focused so far on the ability to distinguish animacy cues from objecthood cues, assuming that this distinction - resulting in significantly different observable patterns - is one of the main sources of infants' early knowledge about animate beings as opposed to inanimate entities. Yet little is known about how the animate vs. inanimate distinction relates to another more fundamental one, that of agents vs. objects. As the authors of this study (Kuhlmeier et al. 2004) themselves explicitly acknowledge, it is debatable how empirical data should be interpreted with respect to this distinction :

It [...] remains an open question whether the results of the present study are due to a distinction between animate versus inanimate entities, intentional agents versus non-intentional objects, or humans versus other entities. (*ibid.*, p.7)

If on the one hand, infants' perceptual representation of objects has been largely studied, little is known, on the other hand, about infants' perceptual representation of *agents*. Spelke [Spelke, 1994] has suggested a number of principles circumscribing the notion of perceptual objecthood in infants. These principles can be formulated, following Bloom [2004, p.12], as :

1. *Cohesion*. Objects are connected masses of stuff that move as a whole. If you want to know where the boundaries of an object are, an easy test is to grab some portion of stuff and pull—what comes with what you are pulling belongs to the same object ; what remains does not.
2. *Solidity*. Objects are not easily permeable by other objects ; if you tap at an object with your finger, your finger does not penetrate.
3. *Continuity*. Objects move in continuous paths ; they travel through space without gaps. An object would violate this rule if it disappeared from one location and reappeared in another.
4. *Contact*. Objects move through contact. A ball on a pool table is not going to move unless something contacts it ; it

will not run from the cue or come when it is called. The exceptions to this rule are animate creatures, like people and dogs, and also certain complex artifacts, such as robots and cars.

Taken together these principles define what counts as an ‘object’ for children. More precisely, they characterize a cluster of properties that children systematically privilege in picking out a certain kind of entities in the environment, prior to any conceptual identification. For this reason such clusters (often referred to as ‘Spelke Objects’) have been described as *proto-objects*.

An analogous characterization of what counts as a perceptual ‘agent’ for infants has not yet been clearly proposed. We argue that a principled inquiry into infants’ ability to distinguish objects from agents cannot be based on mere sensitivity to animacy vs. inanimacy cues, since the ability to perceptually individuate and track agents relies on a much more fundamental distinction : the ability to individuate and track *bearers of intentionality* as opposed to objects.

Are there, beyond mere animacy sensitivity, perceptual mechanisms for tracking agency? To which extent can agents (as opposed to objects) be perceptually individuated, parsed and tracked by children?

We submit that understanding the infant’s perceptual representation of agents might benefit from introducing the notion of a *proto-agent*, i.e. a cluster of properties that determine the way in which infants single out intentional entities prior to conceptual identification.

In order to achieve this aim, we will propose an extension of the object individuation paradigm that relies on Pylyshyn’s proposal for visual object tracking. To construe a notion of *proto-agent*, we will assume :

- (a) following [Blaser et al. \[2000\]](#) and [Pylyshyn \[2001\]](#), that human beings individuate and track single entities via a nonconceptual mechanism which can be operationalized by appealing to the notion of an *object file* ;
- (b) following [Carey and Xu \[2001\]](#), that the same nonconceptual mechanisms underly infants’ early understanding of objects and adults’ ability to keep track of perceptual entities ;

- (c) according to the evidence mentioned above, that there are plausibly different processes that allow children to individuate agents as opposed to objects.

Our main contention is that while tracking perceptual entities endowed with agency (or *proto-agents*) and their persistence over time, the human cognitive system opens a special sort of file, which we will call an *agent file*. Furthermore, we will suggest that the same mechanisms underlie infants individuation of agents and adults' ability to perceptually keep track of agents before explicit identification.

10.2 Perceptual Tracking of Individuals

Nonconceptual mechanisms for individuating and tracking objects have been largely studied in the perceptual literature. In particular, humans' abilities to visually track individual entities have been extensively studied in recent years by Pylyshyn and collaborators [Pylyshyn \[2000\]](#), [Blaser et al. \[2000\]](#), [Pylyshyn \[2001\]](#), [Scholl et al. \[1999, 2001\]](#).

There is robust evidence that a referential link to a particular entity, an individual, is established in virtue of low-level indexing mechanisms (which Pylyshyn dubs 'FINSTs') that :

- precede the deployment of focal attention ;
- operate before and independently from the perceptual identification of the tracked item ;
- are pretty insensitive to major featural modifications in the visual appearance of the tracked items ;

These three aspects suggest the *nonconceptual nature* of such referential mechanisms : they allow a subject to keep track of an item without necessarily encoding specific features or attributes of that item. The nonconceptual nature of this referential link has to be considered, according to defendants of the FINSTs theory, as a precondition to any form of identification or perceptual judgment about objects. In this respect, the visual index theory can be considered as an extension and partial amendment of the hypothesis according to which humans access and store information about perceptual objects in terms of 'files'.

10.2.1 Object Files and Their Dynamics

The notion of an object file is due to the work of [Kahneman and Treisman \[1992\]](#). They proposed that object perception is mediated by the opening of episodic 'files' within which object tokens are constructed. Information about particular perceptual items is thus selected from the sensory array, integrated over time, and stored in such files.

Pylyshyn has suggested that object files can account for the way in which perceptual information is *stored*, but not for the *creation* and *maintenance* of the referential link to the object. Pylyshyn's model enriches and extends the previous theory at two different levels : first, it introduces a mechanism of nonconceptual reference as a requirement for any theory of perceptual reference to individuals ; second, it embeds the traditional object file notion in a larger framework that accounts for the whole dynamics of perceptual items. Following Pylyshyn, hence, we should distinguish three independent classes of properties that are relevant for understanding the dynamics of object tracking, namely :

- (1) INDEX-GRABBING FEATURES
properties that cause the assignment of an index (and enable an object file to be opened for the indexed item) ;
- (2) INDEX-PRESERVATION FEATURES
properties that allow the indexed item to be tracked (and the file to persist over time) ;
- (3) ENCODED FEATURES
properties that can be ascribed to the indexed item (information that can be stored in the file).

Taken together, (1)-(3) characterize how 'files' work : the *content* of a file, i.e. information attached to the indexed individual, is defined by properties of type (3) ; the *dynamics* of the tracking is defined by properties of type (1) and (2). It should be noted that properties belonging to (1) and (2) need not be encoded in the file : they need not be used for identifying entities *as* objects, i.e. they do not provide *per se* any basis for identifying the object or making perceptual judgments about it. Properties belonging to (3), on the contrary, are properties stored in the file, and they can be used for further qualification (categorization/identification) of the object at a conceptual level.

10.2.2 From Object Files to Infants' Representation of Objects

Up to now we have been referring to perceptual objects, i.e. individuals picked out by our perceptual systems without benefit from concepts or prior knowledge. Carey and Xu [2001] have argued that adults' nonconceptual representation of perceptual items and infants' object representation share a number of characteristics :

- they privilege spatiotemporal information in decisions of individuation ;
- they are subject to the same set of size limitations for parallel individuation ;
- they survive occlusion and are sensitive to the distinction between cessation of existence and temporary loss of visual contact (*ibid.*, p.186).

In particular, the individuation mechanisms described in developmental psychology literature and those described by the object tracking literature both seem to privilege entities that are bounded, cohesive and that persist over time.

According to these authors, these strong similarities suggest that the account provided by studies regarding how human adults track perceptual objects and the widely discussed results in developmental psychology regarding how infants represent objects are two descriptions of the *same* system. If their proposal is correct, then it is plausible to assume that the same mechanisms that are put to work when adults track perceptual objects are working when infants track objectual entities : nonconceptual representations of individual entities involved in both domains can hence be described as the same kind, which we refer to using the notion of a *proto-object* [Pylyshyn, 2004c, p.23].

10.2.2.1 Proto-objects vs. commonsense objects

It should be noted that the notion of nonconceptual representation of objectual entities (proto-objects) need not be relevant for our commonsense understanding of *physical objects*. While Pylyshyn [2004c] acknowledges that trackable individual items are typically the proximal counterpart of commonsense physical objects, it is debatable whether this is always the case. We should hence be prudent not to conflate the theoretical notion of proto-

object which is relevant for perceptual and developmental psychology with our shared intuitions about objects, which can be shaped by conceptual, linguistic and cultural factors [Casati, 2004].

The legitimacy of the distinction between a theoretical notion of (proto)objecthood and the commonsense notion of an object motivates our claim that the study of *agency* and agent individuation should respect a similar prescription : what we intend to suggest is that - no matter what our shared intuitions are about agents in everyday life, how we individuate them or identify them - there might be a theoretical notion, comparable to the notion of a proto-object, (we might call it the notion of a ‘proto-agent’) that is relevant for the understanding of nonconceptual individuation and tracking of entities endowed with agency and intentionality. Such a notion should, then, not be conflated with its commonsense counterpart : from now on we will refer to ‘agents’ *tout court* for this theoretical notion of a ‘proto-agent’.

10.3 Perceptual Tracking of Agency

Picking out and maintaining a referential link to agents seems *prima facie* to raise the very same problems involved in referring to objects. We argue that merely being sensitive to agency cues is not sufficient to account for the way agents are individuated, tracked and referred to. What is needed, much as in the case of object tracking, is to understand how a subject is able not only to detect agency, but to maintain reference to an individual which :

- is unique in spite of multiple agency cues ;
- can persist over time ;
- can survive to changes in some of its features ;
- can cease to exist, split or merge with other entities ;
- can be tracked in parallel and independently of other entities of the same kind.

The need for such a mechanism that allows agent tracking can be illustrated through a case like the following :

PLAYING ‘FOOTSIE’. *John is invited for dinner at a friend’s place. He is seated across from a number of guests. At a certain point in the dinner he notices that something is going on under the table : he detects an agency cue, like someone willing to play ‘footsie’ with him. After a while, a second agency cue is detected.*



FIG. 10.1 – Playing ‘footsie’ and agent individuation

John has a problem understanding whether there actually is an agent behind the detected cues, and if it there is, whether one and the same agent is responsible for both cues or more than one agent is involved.

Thus, keeping track of agents seems to require some sort of mechanism for the selection of individuals, the creation of a referential link and its maintenance over time. It is unclear, though, whether and how keeping track of agents could be done using the mechanism used for ordinary object tracking. The mechanism involved in object tracking is not triggered by properties such as purposefulness which could plausibly be a necessary condition for parsing an agent as persistent. If this is so, then such mechanism would not detect the agent’s willingness of playing footsie with John and, thus, would not allow him to keep track of the agent behind the agency cue (not to say that it would be helpless for John’s understanding of whether one and the same agent is responsible for both, the first and the second, cues). Furthermore, there may be cases in which objectual cues conflict with agency cues and it is hard to establish whether individual entities are tracked in virtue of the former or the latter. The literature on infants’ perception of intentionality in self-propelling shapes [Heider and Simmel, 1944, Premack, 1990, Scholl and Tremoulet, 2000] represents a good case study for investigating such issues. Heider and Simmel (1944), for instance, showed that people tend to associate particular intentions or intentionality (doing some good to someone, being good, etc.) to certain systematic ways in which some geometrical figures (circles, triangles, etc.) move while they “tell a story” in a movie. One might ask under which conditions criteria of individuation and

perceptual tracking are based on objectual cues rather than agency cues. It is an interesting empirical issue to study whether :

- an item preserving its agency cues and changing its objectual features can still be tracked as the same agent (*e.g.*, an ‘evil’ triangle turning into an ‘evil’ square without violation of other constraints or an ‘evil’ triangle disappearing and reappearing at another place with a violation of spatio-temporal continuity);
- an item preserving its objectual features and changing its agency features can be still tracked as the same object (*e.g.* an ‘evil’ triangle suddenly turning into a ‘good’ triangle).

Evidence from such experiments supports the claim that there *are* specific individuation processes that depend on agency which can be dissociated from individuation of the same entities as objects [Bloom, 1996]. In what follows we will argue that the question of whether children are able to individuate and refer to agents by appealing to mechanisms similar to those of object tracking might benefit from an extension of the notion of ‘file’ to the case of intentional agents.

10.3.1 Extending the File Notion to Agents

An interesting way to operationalize perceptual individuation and the tracking of agents involves an appeal to the notion of file described above, extending it from the domain of (proto-)objects to the domain of (proto-)agents by postulating the notion of an *agent file*.

If the ability to track agents is not reducible to mere sensitivity to agency cues, in virtue of the agent’s continuity and persistency over time, it seems reasonable to assume that in order to track such entities and ascribe properties to them a subject might use files. Their dynamics can be then described by three classes of properties analogous to those we introduced for object tracking.

1. AGENCY-GRABBING PROPERTIES

properties that enable individuation of an agent (i.e., properties that are necessary to open an agent file);

2. AGENCY-PRESERVING PROPERTIES

properties that allow an individual agent to be tracked (and the file to persist over time);

3. ENCODED PROPERTIES

properties that can be stored and retrieved from the agent file.

In the example introduced earlier, such properties could be : John's detection of an animacy cue on his leg (an agency-grabbing property); the spatio-temporal coherence of this stimulation (for instance its spatial orientation or its frequency : agency-preservation features); any other information that can be ascribed to the agent (including non-perceptual information such as beliefs about the agent's explicit intention of playing 'footsie' : encoded properties).

The main idea is that files are domain-specific so that the kind of information and properties that an agent file can store is different from that which is stored in an object file : the encoded information in the former type of file can include intentions, which can be retrieved in reasoning or while making judgments about the agent.

10.3.1.1 Agent Files vs. Agency Cues

It should be noted that by positing the existence of a class of properties that allow an individual agent to be picked out we are not endorsing the idea that as soon as such properties are *detected* an agent file is automatically opened.

In the case of objects, it has been shown that detection of objectual cues is not per se sufficient for tracking an entity : well-formed perceptual objects might still lack conditions for being tracked over time [Scholl et al., 2001]. The mere presence of objectual cues does not entail the presence of reference fixation cues.

Similarly, there can be cases in which merely detecting an animacy or agency cue, although sufficient to respond to its presence, does not imply that an agent is individuated and susceptible of perceptual tracking. John can for example withdraw his leg as a reaction to the stimulation, without necessarily individuating an agent. Since agent *tracking* is not required in such cases (either because there is no need to individuate the source of the cue or because there is no need to represent this entity as persistent over

time), we assume that opening an agent file is not needed. Actually, we claim that agency-grabbing properties are necessary conditions for agent tracking but they are not *per se* sufficient for the creation of an agent file. Distinguishing between simple detection of properties and full-blooded perceptual tracking - as in the case of detection of objectual cues vs. object tracking - is a natural consequence of our approach.

10.4 Object vs. Agent Tracking : Empirical Directions

Introducing the notion of an agent file raises the problem of understanding whether and how such files are related to files that apply to tracking of objects. If we accept the hypothesis that there are two distinct mechanisms for agent vs. object tracking, it is reasonable to ask how such mechanisms can interact or be mutually related in specific experimental conditions. Broadly speaking, there can be two general options :

(A) INDEPENDENCE VIEW

At each level of description, object files and agent files *share no* common features. Properties relevant for describing the two kinds of files and their dynamics (file fixation, preservation and content) are *necessary and sufficient* to account for the two distinct kinds of tracking.

(B) DEPENDENCE VIEW

At some level of description, object files and agent files *might share* common features. Properties relevant for describing the two kinds of files and their dynamics are *sufficient but not necessary* to account for the two distinct kinds of tracking. Some properties of one kind of file can be exploited at a given level by the other kind of file.

We will outline in what follows some cases in which the relation between the two kinds of files can be empirically studied at each level.

10.4.1 Object vs. Agent File Fixation Cues

Once we acknowledge that object vs. agent tracking can be articulated at three distinct levels, we can raise the question of whether, in the case of

tracking¹, different classes of properties allow *fixing of* reference to an item in the case of agents and in the case of objects. Let us consider an example drawn from a classic arcade game of the '80.

THE ASTEROIDS GAME. *A player must drive a space vessel in order to avoid and destroy both asteroids and enemy vessels. Asteroids are characterized by passive physical movements, while enemy vessels are characterized by motor patterns revealing intentional behavior (like avoiding asteroids, actively changing speed and direction, and shooting other vessels). The player must be able to track both asteroids and enemy vessels and react appropriately to their movements in order to destroy them, survive and win the game.*

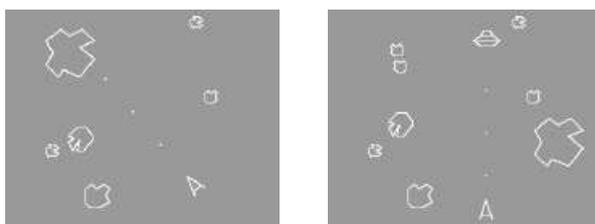


FIG. 10.2 – The Asteroids Game : objects vs. agents

This example illustrates a case in which a subject needs to pick out and maintain reference to two different types of individual (objectual entities, like *asteroids*, vs. intentional entities, like *enemy vessels*). We assume that in order to do this, the subject must detect two classes of cues prior to any further identification or categorization. It seems that, to establish reference, the subject does not need to access properties that might be used to *identify* objects and agents (e.g., asteroids and enemy targets might have the same *shape*). In short, we are claiming that a subject becomes able to individuate and maintain reference to entities belonging to two distinct classes (proto-objects vs. proto-agents) in virtue of his ability to detect objectual vs. agency cues, and of the specific task requirements that force him to maintain perceptual reference to individuals over time. If it can be empirically demonstrated that in similar conditions subjects display robust

¹We do not intend to discuss here cases of mere *sensitivity* to objectual cues vs. agency cues. See the paragraph in section 10.3.1.1 on the distinction between detecting and tracking agency

capabilities to differentially detect reference-grabbing properties belonging to two mutually exclusive classes as a condition for tracking items, then we might plausibly conclude that two distinct and independent individuation processes are at work. This distinction would support the claim that object tracking and agent tracking are *independent* at the level of reference fixation properties.

The existence of two distinct classes of reference fixing properties — although sufficient to support the *independence view* at this level (*agency cues can be segregated from objectual cues*) — is not sufficient to conclude that the two mechanisms of agent and object tracking are necessarily independent. We must also consider the relation between agents and objects at two other levels, viz. reference preservation and feature encoding.

10.4.2 Object vs. Agent File Preserving Properties

We have insisted on the fact that fixing perceptual reference is still not enough for *tracking* individual entities over time. Tracking implies *preserving* a referential link to a perceptual item already picked out. This raises the question of determining in virtue of which properties reference to a single individual can be maintained over time.

In the example above, in order to be able to avoid an asteroid, a subject must be able to track it as *persisting* over time. Following Pylyshyn, we assume that the properties used for fixing reference when the individual is picked out *need not* be the same as those that preserve reference. Once an item is individuated, reference fixing properties can be discarded without breaking up the referential link itself. Provided there are some *file preserving properties*, an item can undergo major changes without losing its singularity : we argue that if preservation conditions are met, an individual will not cease to be treated as a single perceptual item even if the properties initially used for its individuation have disappeared. Assuming that an asteroid was picked out as an individual object through its shape, it is not necessary that its shape be maintained over time in order for the object to persist in a perceptual tracking task.

Recent literature on *Multiple Object Tracking* has demonstrated that tracked items can survive several kinds of disruption of their features. It seems, though, that certain properties are required for an item to preserve its individuality. It is reasonable to assume, on the basis of this literature, that in

our asteroids game example, although asteroids might ‘survive’ temporary occlusions which do not alter their trajectory, they would fail to maintain a perceptual link in cases of spatio-temporal incoherence, like sudden shrinking or disappearing and reappearing at a different location [Scholl and Pylyshyn, 1999]. *Coherence of trajectory* as well as *cohesiveness* [Van Marle and Scholl, 2003] are hence examples of properties that seem to be required in order to maintain reference to perceptual items in tracking tasks.

We might then ask whether the properties used to *keep reference alive* are different in the case of agent or object tracking.

10.4.2.1 An Objectual Bias in Agent Files Preservation ?

Many observable properties might in principle be recruited to preserve agency. For instance, the reiteration of animacy cues (e.g. *an enemy vessel periodically shooting at the player*) or the lack of cues of agency disruption (e.g. *the lack of observable indicators of an enemy vessel being destroyed*) are good candidates for the properties that contribute to the preservation of agent files. This might suggest that at the level of file preservation as well, agents and objects do not share any property (and hence that *object files are independent from agent files*). Nevertheless, given the fact that intentional entities are *usually* coinstantiated with objectual entities (‘bodies’) in our environment, it is plausible to assume that agents can be perceptually tracked via objectual preservation properties. We call this phenomenon an *objectual bias* in agent files preservation. This amounts to saying that :

- (a) there is a class of properties that are sufficient to maintain reference to an object, once an object file is opened.
- (b) there is another class of properties that are sufficient to maintain reference to an agent, once an agent file is opened ;
- (c) among each of the above classes there are single properties that are more or less strongly correlated with preservation of the file : in the case of objects, there might be properties strongly or weakly correlated with object persistence ;
- (d) one and the same property can be correlated with different degrees of reliability to agency or objecthood preservation ;

- (e) object preservation properties that are weakly correlated with preservation of agency might nonetheless be recruited for preserving agency when other agency preservation cues are absent or not available.

The fact that one and the same property P might be relevant to different degrees for maintaining reference either to an object or to an agent does not threaten *per se* the very possibility that there are two distinct kinds of file. We are just suggesting that it is empirically possible that, in virtue of the robustness of the agent-body correlation in our environment, nonconceptual tracking of agents might exploit objectual properties.

Which classes of properties are actually being used for tracking a proto-agent (whether they are properly described as agency-related rather than object-related) is, thus, an empirical issue that calls for experimental work. We maintain that — until a number of explicit conditions for distinguishing between these two classes are formulated — empirical results supporting the hypothesis that infants track individual entities as *persistent objects*, as *persistent agents* or as *agents persisting in virtue of their objectual features* might be seriously undermined. As a possible suggestion, one might test whether the fact of using objectual properties for keeping track of agents' persistence has any consequences on the number of items an infant is able to individuate and track at the same time.

10.4.3 Object vs. Agent File Content

Once the conditions for fixing and preserving an item are met, we have at our disposal a device (a file) to *store* information attached to this individual. The way in which a file attaches properties to an entity is radically different from the way in which the nonconceptual mechanism at work while tracking an individual establishes and preserves reference to it.

On the one hand, nonconceptual tracking provides a *direct link* to an individual, i.e. allows a particular entity to be grabbed and segregated from other individuals or other properties of the scene : tracking is thus a necessary condition for parsing an individual as persisting.

On the other hand, having a file attached to an individual is required in order to ascribe some properties to it.

In the object file literature, a file *content* typically consists of properties that

can be *predicated* of an object and used in perceptual judgment, categorization and identification. Following Scholl et al. [1999, p.2] :

[This kind] of property determines the object's appearance — what a particular object looks like — including its color, shape, lightness, and texture. We call these *featural properties*.

If our hypothesis on the extensibility of the object file paradigm to the case of agent tracking is valid, then we can think of the content of an agent file as a temporary structure attached to an individual entity which can store information about this intentional entity.

It is an empirical issue to understand whether and what are the particular constraints on properties that can be stored in an agent file as opposed to an object file. Yet, we submit that the extension of the file notion to the case of agency can explain how perceptual agency properties can be *ascribed* to a particular agent and used in perceptual judgment.

When an infant is asked to identify among a number of perceptually available agents *which is* the agent displaying, say, aggressive behavior, we claim that she is making use of information stored in a file to perform this task. It should be noted that since properties stored in a file are those properties that are conceptually accessible, they need not have any relevance for understanding how the nonconceptual parsing and tracking of individual entities is done.

10.5 Are There Really Agent Tracking Mechanisms ?

The previous paragraphs were meant to outline a number of empirical issues related to the possible distinction of object vs. agent tracking at three different levels : (a) the level of reference-fixing, (b) the level of keeping reference alive and, (c) the level of ascribing properties. It might be objected that our proposal of an extension from the domain of objects to that of perceptual agents is in fact a mere redescription of the functional role of object files and thus the idea that we are able to track agents in virtue of dedicated mechanisms would be brought into question. The proposal underlying such objection can be called the *deflationary view* on agent tracking.

Our reply to this objection can be articulated at different levels.

- The rationale for the existence of agent tracking mechanisms is that if we want to account *not only* for detection of and sensitivity to agency or animacy cues as opposed to objectual cues, but also for the ability to maintain reference to an intentional entity *persisting* over time, then we need to explain how this representation of persistence is achieved.
- From the fact that there might be significant similarities in the dynamics and nature of object vs. agent tracking (what we called the *dependence view*), it does not follow that the latter should be reduced to the former. In the previous paragraph we made some suggestions about possible empirical ways to assess the similarity/difference between the two mechanisms.
- Our proposal is consistent with a large literature in developmental psychology that has demonstrated the existence in children of two distinct domains of perceptual properties : the domain of animacy and the domain of objecthood. Our contribution can be considered as a framework for extending these investigations to the question of how individuals endowed with animacy and agency can be grabbed as entities persisting over time.
- The hypothesis of the existence of agent files and agent-related tracking abilities opens up some interesting research directions leading to the study of possible conflicts and dissociations between object and agent tracking. Empirical research might shed light on the fact that :
 - intentional and objectual entities can compete for the same attentional or computational resources : it is possible that the limit on the number of items a subject can track at one time is dependent on (or independent of) the class of tracked items (*objects only*, *agents only* or *objects plus agents*) ;
 - there may be interesting cases of dissociations, i.e. selective impairments of either of these abilities without functional consequences for the other : we might imagine cases of subjects being able to track objectual entities but not agents or viceversa.
 - inattentive blindness studies might investigate whether the existence of two distinct classes of entities has any effect on their neglect : this might suggest that the traditional list of ‘styles of attention’ (object-based vs. space-based attention) should be extended to include a third kind of style (agent-based attention) ;

- developmental investigations might benefit from the notion of an agent tracking mechanism to establish the conditions under which infants represent the *persistence* of a perceptual item over time.

The above considerations suggest that dedicated mechanisms for agent tracking are likely to have psychological reality and account for a number of capacities involved in perceptual reference to persistent entities endowed with agency. The relation of such mechanisms to those involved in perceptual reference to objects remains, however, an open empirical issue deserving further investigation. We have given arguments in favor of a moderate *dependence view* suggesting that in some cases objectual properties might be recruited to establish and maintain reference to agents.

10.6 Conclusions

A crucial step for understanding our nonconceptual abilities to refer to individual entities consists in explaining how such individuals are *picked out* and *tracked over time*, prior to any form of categorization or conceptualization. Such mechanisms of direct reference to individuals lay probably at the basis of both infants' abilities to parse objects and adults' capacity to keep track of multiple perceptual items. We have proposed an extension of the studies on object individuation and tracking to the domain of perceptual individuation and tracking of entities endowed with agency. This extension is intended to fill a gap between the study of perceptual sensitivity to agency cues and a full-fledged understanding of how perceptual items tracked as agents (what we called *proto-agents*) can persist over time. We have argued that unless some explanation of the mechanisms underlying agent *persistence* is provided, many empirical results concerning infants abilities to track animate entities could be dramatically undermined. Our proposal of three distinct levels that might be involved in agency fixing, preservation and ascription is meant to provide the basic requirements for any explanation of perceptual capabilities to track agents. We argue that by analyzing the dynamics of agents at these three different levels, more principled answers might be given to the question of possible interferences and biases between mechanisms dedicated to agent vs. object individuation.

Chapitre 11

Percept

The Notion of the Non-Conceptual in an Explanation of Illusion

[Article écrit en co-auteur avec Nivedita Gangopadhyay, soumis à la date de la rédaction.]

11.0.1 Abstract

The explanation of the classic case of the Müller-Lyer illusion necessitates the postulation of a structured non-conceptual level. Within the domain of the non-conceptual we propose a theoretical entity that we call a “percept” which composes what we call a “proto-judgment” and we present a number of characteristics that we argue proto-judgments possess viz. being non-conceptual, non-epistemic, non-representational entities composed of percepts and which determine the cognitive agent’s behavior with regard to the content of her experience. We claim that the postulation of such a non-conceptual entity helps provide a satisfactory explanation of the Müller-Lyer illusion. We conclude by proposing a three-level model of experience that is needed to explain perceptual phenomena such as the Müller-Lyer illusion.

11.1 Introduction

In a recent paper E. Pacherie (Pacherie, 2000) has suggested that the dichotomy proposed by Dretske (Dretske, 1969) between epistemic (cognitive) and non-epistemic (simple) seeing may be an oversimplification of

the whole story of perception. Pacherie argues that there are instances of perceptual experience that do not allow of a clear classification into either of these two types of seeing and that a satisfactory account of perceptual phenomenon requires the postulation of a third level of perceptual content intermediate between the epistemic and non-epistemic levels. This hypothesized level, a structured non-conceptual level, is invoked to explain among others the problem of illusion. In case of illusion, whether they be cognitive or perceptual, there is a major discrepancy between the actual state of affairs in the world and our perception of it and hence an analysis of an instance of an illusion may help us determine the extent of influences of the sensory data coming from the world on the one hand and our own mind on the other on the resulting percept. We will call the hypothesized non-conceptual entity a “proto-judgment”. We will argue that a proto-judgment is a structured non-conceptual, non-epistemic, non-representational entity composed of percepts, i.e. by invariants of perception, which determine the cognitive agent’s behavior with regard to the content of her experience. It determines the behavior of the agent in that it is a persistent impression of an experience following which the agent will act. Protojudgments may appear in all cases of perception but it is easier to account for their possible existence in cases of illusions such as the Müller-Lyer illusion. Before we take up the classic case of the Müller-Lyer illusion for analyzing it in the light of the proposed non-conceptual level of content we would like to present a brief characterization of what we mean here by conceptual and non-conceptual.

11.2 The conceptual and non-conceptual contents

The distinction between conceptual and non-conceptual content was explicitly put forward for the first time by G. Evans (Evans, 1982) who maintained that information obtained from the perceptual systems is initially non-conceptual (and unconscious) becoming conceptual (and conscious) only when used in thinking and reasoning. Later writers on the subject (Crane, 1992, Peacocke, 1989,1992,1998, Cussins, 1990, Bermudez, 1995, Hurley, 1998, Tye, 1995) have variously characterized non-conceptual content arguing for the most part that the richly detailed and fine-grained nature of perceptual content cannot be accounted for only in terms of conceptual ca-

pacities. In this paper by “concept” we will accept the definition proposed by Carey, 2004.

A concept is a unit of thought which is a constituent of larger mental structures as belief, judgments, desires and so on. Concept is representational. It is individuated by its reference and its role in inferential structure (Carey, 2004, p. 60). Concepts can involve high level abstraction and generalization yielding mental entities that are indeed so far removed from what is ordinarily given in sense-experience that the semantic content of these concepts is almost always confined to the realm of the mental (e.g. concepts of God, perfection, truth etc.). Concepts can also belong to lower levels of abstraction and generalization resulting in mental entities that act as a general term for things in the world that can be said to be tokens of that concept (e.g. concept of table, red, etc.). However, irrespective of whether concepts belong to a high level or to a low-level they share the feature of being structured. They can be thought of as inter-linked and they allow inferences. This feature leads to the further implication that the concepts that an agent possesses are consistent among themselves in the case of a given experience. This consistency of concepts makes it theoretically impossible to account for perceptual phenomena such as the Müller-Lyer illusion in terms of concepts as the explanation would have to involve ascribing contradictory concepts about the same experience. We will take up this line of thought in Sec 3. while discussing the role of non-conceptual content in these cases.

On the other hand, non-conceptual content of experience is that content of experience which an agent can possess without having the concepts that characterize the content. However, it is to be noted that the term “non-conceptual” can be applied to a state as well as to the content of a state. When used to qualify a state it refers to a state which is not brought forth by the conceptual capacities of the agent or a state which cannot be accounted for by the concepts possessed by the agent. “Non-conceptual content, on the other hand,” can be understood in two ways. First it can refer to those contents of experience in order to have which the agent need not possess the concepts required for describing this content. This is not to say that the agent as a matter of fact does not possess those concepts but rather that the concepts play no role in determining these contents of experience. This is to highlight that an agent can experience things for which she has no concepts. Secondly, non-conceptual content can also be understood in a

negative way i.e. as that which is not conceptual. Concepts constitute high level mental entities that enable the agent possessing them to categorize, make abstractions and inferences and for this concepts must satisfy the Generality Constraint. The Generality Constraint is stated by Evans as- “Thus, if a subject can be credited with the thought that a is F , then he must have the conceptual resources for entertaining the thought that a is G , for every property of being G of which he has a conception. This is the condition I call ‘The Generality Constraint’ ”. (Evans, 1982, p. 104). So when construed negatively as that which is not conceptual, non-conceptual content is that which does not permit the same kind of categorization, abstraction and inferences as concepts and which does not satisfy the Generality Constraint. In this paper we seek to provide a positive categorization of non-conceptual content. Whether or not conceptual and non-conceptual contents are completely different with respect to their features remains to be discussed and our aim in this paper is to show that they may not be absolutely distinct. We will argue that cases of illusion such as the Müller-Lyer illusion indicate that the non-conceptual admits of a structure that is needed for explaining the illusion and this entity seems to possess some features similar to ones that concepts have yet radically differ from the latter by the lack of certain crucial aspects of these features.

11.2.1 The epistemic and non-epistemic levels

In order to bring out the implications of Pacherie’s (Pacherie, 2000) claim that the epistemic and non-epistemic dichotomy proposed by Dretske (Dretske, 1969) does not account for a number of perceptual phenomena, like for instance the Müller-Lyer illusion, there is the need to clarify the possible relationships that can obtain between conceptual and non-conceptual content and epistemic and non-epistemic levels. Dretske (Dreske, 1969) explains the notion of epistemic seeing by enumerating a number of conditions that a perceptual state must fulfill in order to qualify as an instance of epistemic seeing.

First, S sees that b is P in a primary epistemic way only if :

(i) b is P

Secondly, S sees that b is P in a primary epistemic fashion only if :

(ii) S sees b

Thirdly, S sees that b is P in a primary epistemic way only if :

(iii) The conditions under which *S* sees, *b* are such that *b* would not look, *L*, the way it now looks to *S* unless it was *P*.

Finally, *S* sees that *b* is *P* in a primary epistemic way only if :

(iv) *S*, believing the conditions are as described in (iii), takes *b* to be *P*.

With regard to the notion of non-epistemic seeing Dretske states, "There is a primitive visual ability which is common to a great variety of sentient beings, an ability which we, as human beings, share with our cocker spaniel and pet cat. It is an endowment which is relatively free from the influences of education, past experience, linguistic sophistication, and conceptual dexterity." (Dretske, 1969, pg.4). In this context it is important to note that the epistemic and non-epistemic levels suggested by Dretske are logically distinct and need not be psychologically independent. Thus Dretske claims, "It may be that adult human beings, by virtue of their past experience, cannot see a candle-stick holder, under any conditions, without at least being brought to the belief that they are visually aware of something. But this fact, if it is a fact (and I do not think it is), would not tend to show that it is logically inconsistent to say of someone that he saw a candle-stick holder without believing himself visually aware of anything, that seeing a candle-stick holder involves, as part of its meaning, the percipient's believing something of this sort." (Dretske, 1969, pg.10). From such claims of logical independence between epistemic and non-epistemic seeing it is possible to consider the notion of non-epistemic in two ways, viz.

i) An agent can possess a mental state without being aware that she possesses it

and/or

ii) An agent can possess a mental state without knowing why or how she came to have it. This amounts to saying that an agent can possess a mental state without possessing the justifications for the state.

So there are two ways in which a mental state could be considered as being non-epistemic. The first is that the state has an unconscious content or a content of which the agent possessing the state is not aware, i.e. the agent cannot grasp the contents of the state, and secondly the agent is conscious or aware of the content, i.e. she has an access to the contents or can grasp them, but does not know the reasons as to why she possesses them. From this we can derive the features of a state at the epistemic level too. A state would be epistemic if

- i) The agent possessing the state is aware that she possesses it, i.e. the agent can grasp the content of the state. An example of this kind of state can be a belief state where the agent is aware that she possesses the belief but can offer no reason for it

and/or

- ii) The agent possessing the state knows why and how she came to have it. In other words the agent has the state as well as the justification for it. This feature would promote the state in question to the status of knowledge. By “knowledge” here we accept the classical view of justified true beliefs.

Let us now consider the possible relationships that could hold between conceptual and non-conceptual content and epistemic and non-epistemic levels. A state can have a conceptual content and belong to the epistemic level by satisfying both or at least one of the above-mentioned features of the epistemic level. A person possessing a state with a conceptual content is necessarily aware that she possesses the concept and she may also be knowing why and how she came to possess it. If the agent merely possesses the concept without having the justifications for it, it is of the nature of a belief which may be true or false. For example, an agent may possess the concept of God but if questioned may not be able to provide justifications for possessing the concept. Here the agent has a mere belief in the concept and not knowledge. But if she also knows the reasons for possessing the concept and if the belief is true then it becomes knowledge. For example, a person has a concept of snow and justifies the possession of the concept by saying that every time certain climatic conditions are satisfied the water droplets in the clouds will fall as snow. In this case we can say that the agent knows what snow is. However, the possession of a concept is not necessarily linked to the possession of the justifications for the concept and an agent can have a concept without the justifications. Thus the following relationships are possible between a conceptual content and the epistemic and non-epistemic levels :

A conceptual content necessarily belongs to the epistemic level if we consider the first feature of the epistemic state i.e. being aware. However if we take into account the second feature then a conceptual content may or may not be epistemic. From this the relationships between a conceptual

content and the non-epistemic level can also be derived :

A conceptual content can never belong to a non-epistemic level in the first way. We cannot both have a concept and not be aware that we possess it. However, since merely possessing a concept does not amount to knowledge, a conceptual content can be non-epistemic in the second sense. An agent can possess a concept X without knowing why or how she has it or without being able to give any condition of use and application of this concept.

The picture is considerably different with respect to non-conceptual content. A non-conceptual content can belong to the epistemic level in the first way but it can never belong to the epistemic level in the second sense i.e. an agent cannot offer any justifications as to why she has a particular non-conceptual content even if she is aware of having the content. A non-conceptual content can, however, belong to the non-epistemic level in both ways. There can be non-conceptual content of which we are not aware, e.g. sensory data, and there can also be non-conceptual content which we possess but do not know why and how we do so. For example, in the Müller-Lyer illusion we have the persistent perceptual impression that the lines are not of equal length but we do not know why and how we see them as unequal.

11.2.2 The representational aspect

Since the distinction between conceptual and non-conceptual content is generally drawn within a representational account of experience, it is important to specify the possible relationships between conceptual and non-conceptual contents and representation. Without delving into the intricate and baffling question of what is a representation, we will accept a notion of representation as simply that where a thing stands for something else. What is of immediate interest to us is the idea of correctness associated to the notion of representation. Peacocke argues that a mental entity can be called representational if it is subject to the correctness condition (Peacocke, 1983). An entity that represents something can be said to be a symbol for the thing it stands for in a system that processes symbols. The fact that an entity represents another entity is determined by a logical function governing the input of objects and the output of values. Thus, α can be said to represent a if $f : \alpha \mapsto a = 1$. In the context of a bivalent logic can have two values, $\{0, 1\}$.

In order to apply this function it must be verifiable. It must be possible to state that $f : \alpha \mapsto a = 1$. Consequently, there must be a function \mathcal{F} which allows us to determine if the function f determining the value of 1 for $\alpha \mapsto a$ is correct or not. This function \mathcal{F} can be called the evaluation function or the correctness condition. Hence this correctness condition can be said to be sort of in-built in the notion of representation. Thus if the agent possessing a mental entity can revise it or change it or correct it the entity can be said to be representational. It is this condition that enables us to talk about “errors” and “mistakes” on the part of the cognitive agent. Now in case of conceptual content it can be said that these contents meet the correctness requirement. Perhaps the most compelling evidence comes from cases of erroneous perception. A person perceives a stick half-immersed in water as bent. As a result of this perceptual experience she gathers the belief - the stick is bent. This is certainly a conceptual state but it is open to revision. After she realizes that the stick is actually straight but was looking bent in water she replaces her former belief by the belief- the stick is straight. Hence her previous conceptual state, that is her belief or judgment, has been modified. Since concepts allow of such correction, they meet the correctness condition and hence can be said to be representational. All concepts satisfy the Generality Constraint and thus can be said to be representational in this sense (Evans, 1982). Non-conceptual content, however, is not subject to the correctness condition. This claim is supported by the analysis of instances of illusions such as the Müller-Lyer illusion. In this case no matter how much we try we cannot but see the lines as unequal. Even after we measure the lines and discover that our perceptual experience of seeing them as unequal is actually an illusion, the percept does not change. The agent cannot change the percept as a result of changes in the belief state concerning the same experience. However, the change of the percept as a result of changes in the environment or the stimulus presented is quite a different issue. After the discovery of the illusion there is a change in the belief-state and hence in the conceptual content but this change does not affect the actual experience. To explain this phenomenon the non-conceptual content is introduced as it can be explained neither by the conceptual content nor by the mere sensory data as we will argue in Sec.3 , following Pacherie. As non-conceptual content does not allow change or revision it cannot be representational in the sense of meeting the correctness criterion.

11.2.3 Proto-judgments

From the above discussions we can enumerate the following characteristics of conceptual and non-conceptual content : Conceptual content can be i) epistemic (in both ways), ii) non-epistemic (in the sense of lacking justifications), iii) representational, iv) subject of consciousness or awareness. In addition to these concepts also have the following crucial features. They are lexical, may or may not be linguistic as a subject can possess a concept without knowing the corresponding general term, have corresponding abstract entities and involve abstraction.

Non-conceptual content – i) cannot be epistemic, ii) is non-epistemic (in both ways), iii) is necessarily non-representational, iv) may or may not be conscious (a perceptual state like the one in case of the Müller-Lyer illusion has a conscious non-conceptual content but sensory data are example of unconscious non-conceptual content). Moreover, non-conceptual entities are lexical (expressible through non-linguistic behavior at least), are non-linguistic, do not have corresponding abstract entities and involve abstraction although not of the same kind as involved in concepts. In this context by abstraction we mean the features of being “context-free” i.e. not bound to one particular context. An entity derived by abstraction would thus be one which can be applied in several contexts. However, the crucial distinction between concepts and non-conceptual entities in this respect is that in addition to involving abstraction in the sense of being “context-free” concepts also satisfy the Generality Constraint whereas non-conceptual entities, such as invariants, are context-free as they can be applied in several contexts but they do not satisfy the Generality Constraint. Thus concepts and non-conceptual entities both involve abstraction although not of the same kind. Concepts can involve different degrees of abstraction and generalization and they can be linguistically expressed by beliefs and judgments. The presence of abstraction and inference makes the conceptual contents structured. In this case it is to be noted that, if we keep aside the controversial claim of innate ideas, concepts are initially formed by abstraction from empirical data. The processes of abstraction operating in the formation of concepts pick out common features of a number of entities sharing some similarity and combine them to build a concept. The greater the abstraction and generalization, the higher is the concept. Low level concepts can be said to be less distant from objects of the world than higher level concepts. It can

also be said that some lower level concepts are nothing but the perceptual representations of entities of the world. Prinz's notion of "proxytypes" could serve as such low level conceptual entity (Prinz, 2002). Proxytypes are used as detection and tracking mechanisms and they can also be incorporated in thought via working memory. These proxytypes are of the nature of primitive perceptual/imagistic representations.

On the other hand, in case of non-conceptual content abstraction and categorization is operative though they are not the same as in case of conceptual content. At the first level of the non-conceptual realm there is the pick-up of information by the sensory organs. This information is obtained in a sense-specific way i.e. the visual sense organs picks up only visual cues in the environment and the same is the case with other senses. However, not all cues available to a sense are relevant to the task at hand. The sense organ will typically pick up salient cues from the environment. As a result of evolution and "attunement" to the environment an organism will typically "resonate" to the invariants that are significant for it (Gibson, 1966, 1977,1979). This implies that among all the cues available, an organism abstracts the salient ones. Invariants can be said to be the salient cues occurring or appearing in several contexts (hence "context-free" as they are not confined to one particular context) but nevertheless being dependent on the context they do not satisfy the Generality Constraint. This pick up of invariants is not conceptual as an agent can pick them up without possessing the corresponding concepts. Neither is this epistemic as an agent can possess or pick up the invariants without knowing that she possesses them or why and how she does so. Neither is this representational as the agent cannot correct the invariants picked up as she is in no position to choose which invariants will be picked up by the senses ; there can be no error at the level of invariants although there could be relevance or irrelevance of the invariants for the task. The fact that an agent can learn to pick up new invariants does not imply that she can correct them. For example, in Pavlov's study of conditioned reflexes the dog learns to salivate at the sound of the bell, i.e. it learns to pick up a new invariant (Pavlov, 1927). But when the newly learnt stimulus is presented without the food, the dog still salivates for a number of trials. This shows that even when food is not presented the dog cannot correct its pick up of the invariant as a salient cue of the environment. When the dog no longer salivates at the sound of the bell if it is not paired with the uncon-

ditional stimulus, it is because it simply becomes an irrelevant cue. Hence it follows that invariants do not share the features of conceptual content but rather fall in the category of non-conceptual content. Nevertheless, this non-conceptual content allows :

- i) Abstraction (picking up of only salient cues in the environment),
- ii) Generalization (the same invariant can be instantiated in several contexts),
- iii) Categorization (the invariant enables the agent to distinguish several features and recognize features of the same invariant kind, here categorization means discrimination)
- iv) Inference (routines can be based on invariant schemes or patterns).

In view of these features of non-conceptual content we can say that this content is also structured. Broadly speaking, the lowest level of non-conceptual content would comprise of those entities which have been the least modified by cognitive activities of the agent. The raw sensory data would in this sense constitute the lowest level of non-conceptual content because at the sensory level the data is still nearly completely uninfluenced by conceptual elements. But higher up we find elements which have been subject to processing by the agent and have undergone modification from the raw data stage. The abstraction and categorization of the raw sensory input provides us with the content of the higher levels of non-conceptual stage. Just as concepts yield beliefs and judgments so does non-conceptual content yield what we will refer to as “proto-judgment”. Proto-judgments occupy perhaps the highest level in the non-conceptual state and just fall short of qualifying as concepts due to their lack of conscious accessibility, linguistic features and resistance to modifications.

Proto-judgments are not conceptual. An agent can possess them without possessing any concepts needed for describing them. They are non-conceptual but not in the way a routine can be said to be non-conceptual. They are non-epistemic in the sense that although the agent can be aware of having a proto-judgment, she does not know why or how she obtained it. A proto-judgment is non-representational as it is not subject to the correctness condition. It cannot be changed, revised or corrected without modifying the environment and the conditions of experience. Correction at the conceptual level has no impact on the proto-judgment. Even after discovering that the two lines of the Müller-Lyer figure are of equal length the agent still has the

strong impression that they are unequal. A proto-judgment is expressible but not linguistically. A so-called perceptual judgment like “The two lines are unequal” would be a description of a proto-judgment at the conceptual level. Proto-judgments are formed through a process of abstraction which can be said to be a “proto-categorization” in the sense that scenario-content (Peacocke, 1992) is a form of categorization of the scene but which being dependent on the object present in the experience does not satisfy the Generality Constraint. Proto-judgments only require discrimination and not identification. Finally, just as judgments are composed of concepts, so proto-judgments are composed of percepts which are abstract invariants of perception; the same percept can appear in several proto-judgments.

11.3 The use of proto-judgments in explaining the Müller-Lyer illusion

Non-conceptual content has been introduced as a theoretical posit to help explain among others the fineness of grain of perceptual experience that goes far beyond the conceptual capacities of the perceiver and also for perceptual phenomena such as the Müller-Lyer illusion that cannot be attributed to either conceptual content or to mere sensory data. In the present context we shall concentrate only on the latter two cases of illusion to analyze the role that non-conceptual content and proto-judgments play in them.

Let us begin by the Müller-Lyer illusion. Figure 1 depicts the classic Müller-Lyer illusion. In this case although the two lines are in fact of equal length they appear to be different in our perceptual experience.

In the case of this illusion our perceptual judgment initially takes the form “The two lines are of unequal length”. After the illusion has been pointed out to us and we have been convinced by taking measurements that the two lines are in fact equal there is a change in our belief state and we can express a judgment of the form “The two lines appear to be unequal although they are actually equal”. There is thus a discrepancy between what the perceiver sees and what she believes. The reason why such an instance of perception cannot be accounted for by epistemic or cognitive perception i.e. by means of concepts is that if concepts determined the content of this experience we would have to ascribe contradictory concepts for a single experience to the perceiver (Pacherie, 2000, Crane, 1992). To explain the

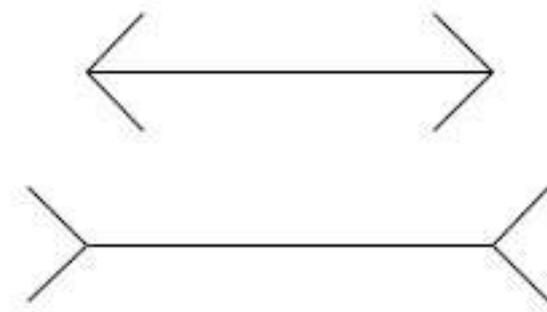


FIG. 11.1 – Müller-Lyer illusion

illusion in terms of cognitive seeing amounts to saying that the perceiver both believes and does not believe that the lines are of equal length. But concepts can be thought of as interlinked and thus they are consistent among themselves within the context of a given experience. An agent cannot have contradictory concepts about the same content of experience. Hence it is not plausible to state that the subject both believes and does not believe that the lines are equal. It is the Principle of Cognitive Significance that “... offers a logical constraint on the content of beliefs.” (Gunther, 2001, pg.187). Thus a more satisfactory explanation of the phenomenon is that the experience has different components - a component which depicts the lines to be of unequal length and a conceptual/belief component that reveals them to be of equal length. Thus a purely conceptual content cannot account for the illusion in question.

However, the perception that the lines are of unequal length cannot be attributed to the bare sensory data (non-epistemic) either (Pacherie, 2000). The illusion cannot be attributed to the state of affairs in the world and their simple pick-up by the senses because the information coded in the diagram is that the two lines are of the same length. The question thus arises- if our raw sensory data does not furnish us with the information that the lines are unequal and if we believe that they are in fact equal then how is it that our final percept or perceptual content is that the lines are unequal? To provide a satisfactory explanation of this phenomenon we will usher in the notion of non-conceptual content.

As the analysis of the Müller-Lyer illusion demonstrates, the content of the perceptual experience or the percept can neither be wholly accounted for in terms of concepts nor in terms of sensory data. Hence this phenomenon is not determined by a conceptual content but neither is it determined by the lowest level non-conceptual content i.e. the raw sensory data. It rather seems plausible to maintain that the percept here is created by a non-conceptual level that has undergone significant abstraction and categorization from the original raw data but which, nevertheless, does not rise to the conceptual level. It seems to embody a high level non-conceptual content, a content that has undergone abstraction from the invariants, which is conscious but regarding which the agent does not know how or why she acquired it, which is not linguistically expressible and which can never be revised or corrected. In short, this is a proto-judgment. Perceptual judgments of the form- The two lines always appear to be unequal- are a description of this proto-judgment at the conceptual level.

In the Müller-Lyer figure had the lines been marked with arrows in a similar way the illusion would never have taken place. The fact that in the figure the arrows point towards opposite directions sort of “distracts” the visual system from forming the correct perceptual judgment about their length from the sensory data. After otherwise determining that the lines are in fact equal the conceptual/belief system revises its first incorrect belief about the length of the lines. But this correction at the level of concepts does not in any way change the fact that the lines continue to appear unequal to perceptual experience. No matter what we do we can never perceive the actual state of affairs in such a context. We cannot make any corrections at the level of percept. Neither can we be held “responsible” for such erroneous perception. The perceptual system seems to possess a component that has been hard-wired in such a way that it will invariantly pick up the cue (or the “distractor” or distracting cue) that hints at the difference in length disregarding other cues that can indicate the reality. Hence we cannot but see the lines as unequal. This feature cannot be attributed to the sense organs or to the raw sensory data as here it is impossible for the raw data to supply the information that the lines are of unequal length simply because they are not so! Hence the flow of information coming to the senses from the scene cannot be charged with creating the illusion. This rich sensory data is selected by further processing to yield the final percept and in the stages of selection

there seems to be a component that is attuned to certain cues among others and presents them as being salient and the resulting percept is thoroughly influenced by such a selection of cues. Yet the claim that this selection does not take place wholly at the conceptual level is bolstered by the fact that we cannot revise and change our percept. We can only change our belief about it. However, although this level cannot be controlled by us (i.e. by our conceptual schemes) it may manifest some functions not unlike the ones found at the conceptual level. The reason for such a claim is that during the selection process from the raw sensory data there must be abstraction and categorization of the incoming information. The abstraction can take the form of picking up of the salient cue from a multitude of inputs and the “proto-categorization” can be grouping them into consistent entities, distinguished from other similar groups, with each group furnishing links between its constituents. The content of such a non-conceptual state is what we term “proto-judgments”. They differ with respect to conceptual content in that the latter is generally linked with linguistic capacities and are open to revision.

Thus the Müller-Lyer illusion demonstrates that the processes of abstraction and a kind of categorization of the raw sensory data have yielded an element which is not exactly what there is out in the world but again which is not completely dissociated from external reality. A certain state of processing in the agent picks up some cues as it is hard-wired to do and passes them on to levels that are manipulated by the agent but which are not influenced by the conceptual level. Proto-judgments thus incorporate how in cases of perceptual phenomena such as the Müller-Lyer illusion the data from the external world and the contributions from the conscious agent interact to produce the phenomenon of perception.

11.3.1 The theoretical levels of experience

The account of the perception of the Müller-Lyer illusion that we have presented so far indicates that at least some, if not all, experience can involve a number of theoretical levels that are logically distinct from each other. At the lowest level is the pick-up of information by the sensory organs. This sensory level which constitutes the lowest strata of the non-conceptual domain is composed of invariants and routines composed of those invariants. These are not available for conscious access by the agent. There is no ques-

tion of any “mistake” as such at this level because the agent can in no way control the pick-up of cues from the environment. The senses have been sort of “programmed” through evolution to pick-up certain cues in particular contexts and to ignore others. If the cues picked up do not contribute to the successful execution of the task at hand they are simply “irrelevant”, if they enable the agent to successfully carry out the task they are relevant. But salient cues can only be irrelevant or relevant; they cannot be “right” or “wrong” as they cannot be revised even if our conceptual system informs us that the information abstracted from the cues is not depicting the actual state of affairs in the world. We cannot evaluate the picked up cues independently of their actual occurrences in the contexts. The evaluation is thus pragmatic and not objective in the sense that they could be right or wrong. The invariants are neither representational entities. They are not symbols standing for some thing else in the world and they are not subject to the correctness condition. They are the inputs for the perceptual system that immediately gives rise to perceptual routines that are goal-directed, mechanical, unconscious, non-epistemic, non-representational and non-linguistic.

Higher up in the non-conceptual domain, the abstraction and proto-categorization of the invariants lead to the percepts and the proto-judgments composed of these percepts. The percept and the proto-judgment are theoretical entities that need to be postulated in order to explain instances of perception such as the Müller-Lyer illusion. A percept is less than a concept as it is not subject to the correctness condition and hence not representational; it is also non-linguistic unlike a concept. However, although non-conceptual, percept is not equivalent to invariants. Invariants are goal-dependent as cues are judged to be salient or not salient with respect to the task to be accomplished. Percepts are not goal-dependent in this way. Moreover, percepts are consciously accessible whereas invariants are not so accessible. This access could be a phenomenological or first-person access or it could be a cognitive access where the agent has access to the personal history of the percept (as could be the case for phobias for instance). Just as invariants yield routines so percepts yield proto-judgments. Proto-judgments cannot be equivalent to judgments at the conceptual level because they lack the linguistic component and the representational (and hence correctional) feature. Yet they are not synonymous with routines precisely because of the same reasons that distinguish a percept from an invariant.

Finally, there is the conceptual level. Concepts and judgments and beliefs formed from these concepts constitute the entities at this level. These entities are epistemic, may or may not be linguistic, are subject to revision and hence are representational. Although in cases of perception as the Müller-Lyer illusion, concepts do not play any role in determining the percept (and hence the proto-judgment), judgments at the conceptual level can at times offer a description of the percept and proto-judgment at the conceptual level. Here it is to be noted that one of the primary reasons as to why the percept in case of the Müller-Lyer illusion could not be attributed to the conceptual level was because of the Principle of Cognitive Significance according to which an agent cannot have contradictory beliefs about the one and the same content of experience. This seems to imply that while principles of rationality are operative at the conceptual level, the non-conceptual level appears to be relatively free of them. However, whether this contention is acceptable or not remains to be seen ; but at least if we concede that at the non-conceptual level some forms of abstraction, categorization and generalization etc. exist it may indicate that non-conceptual content has some laws of its own which may not be of the same kind as the laws of thought at the conceptual level but which, nevertheless, serve to bring order and organization into non-conceptual content.

11.4 Conclusion

This paper has presented a purely theoretical analysis of a perceptual phenomenon such as the Müller-Lyer illusion. We have argued, following Pacherie, that the explanation of this phenomenon requires the postulation of a mental entity called non-conceptual content. We claim that it is this non-conceptual content that constitutes the proto-judgment of the Müller-Lyer illusion. Proto-judgments are non-conceptual, epistemic in the sense that the agent is aware of them, non-linguistic but not non-lexical, and non-representational. However, so far the claim is purely theoretical. Whether proto-judgments can be empirically justified or not remains to be seen. Moreover the claim that an agent can have phenomenological access to proto-judgments along with some other features of proto-judgments such as being non-representational (incorrigible) and non-linguistic could raise the question- are proto-judgments the same as qualia? An initial insight in this

issue would yield a negative response as qualia are supposed to characterize “what it is like to be” in a state X whereas proto-judgments are primarily concerned with the actual content of the perceptual experience and not with what it is like for the subject to be in the state. In case of the Müller-Lyer illusion the actual content of the perception, which is the perception that the lines are unequal, is what is constituted by the proto-judgment without any reference to what it feels like for the agent to be in this state. However, the issue is indeed intricate and requires greater analysis to satisfactorily determine it.

References :

- Bermudez, J.L. (1995). Non-conceptual Content : From Perceptual Experiences to Subpersonal Computational States. *Mind and Language* 10 : 333-69.
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Dedalus*, 113(1) :58–69.
- Crane, T. (1992). The non-conceptual content of experience. In T. Crane (ed.) *The Contents of Experience*. Cambridge, Cambridge University Press, pp. 136- 157.
- Cussins, A. (1990). The Connectionist Construction of Concepts. In *The Philosophy of Artificial Intelligence*, ed. M.A. Boden. Oxford : Oxford University Press.
- Dretske, F. (1969). *Seeing and Knowing*. Chicago : The University of Chicago Press.
- Dretske, F. (1981). *Knowledge and the Flow of Information*, Cambridge, Mass. : MIT Press.
- Dretske, F. (1988). *Explaining Behavior*, Cambridge, Mass. : MIT Press.
- Dretske, F. (1995). *Naturalizing the Mind — The Jean Nicod Lectures - 1994*. Paris : Editions du CNRS & Cambridge, Mass. : MIT Press.
- Evans, G. (1982). *The Varieties of Reference*. Oxford : Clarendon Press.
- Fodor, J, A. (1998) *Concepts, Where Cognitive Sciences Went Wrong*. Oxford : Clarendon Press.

- Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston : Houghton Mifflin.
- Gibson, J.J. (1977). The theory of affordances. In R. Shaw & J. Bransford (eds.), *Perceiving, Acting and Knowing*. Hillsdale, NJ : Erlbaum.
- Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston : Houghton Mifflin.
- Gunther, Y.H. Content, Illusion, Partition. *Philosophical Studies* 102 : 185–202, 2001
- Heil, J. (1991). Perceptual Experience. In B. McLaughlin (ed.), *Dretske and his Critics*. Cambridge, Mass. : Basil Blackwell, pp. 1-16.
- Hurley, S. (1998). *Consciousness in Action*. Cambridge MA : MIT Press.
- Kelly, S. (2001). Demonstrative Concepts and Experience. *Philosophical Review* 110, no. 3 (July 2001) : 397-420.
- Martin, M.G.F. (1994). Perceptual Content. In *A Companion to the Philosophy of Mind*, ed. S. Guttenplan. Oxford : Basil Blackwell.
- McDowell, J. (1994). *Mind and World*. Cambridge, MA : Harvard University Press.
- McDowell. (1998). Replies. *Philosophy and Phenomenological Research* 58 : 403
- Meini, C. (1996). La frontière entre percepts et concepts — La théorie de Dretske à l'épreuve des illusions d'optique. *Incognito*, 6, 27-38.
- Murphy, G. (2002). *The Big Book of Concepts*. The MIT Press.
- Pacherie, E. (2000). Levels of perceptual content. *Philosophical Studies*, 100, 237- 254.
- Pavlov, I.P. (1927). *Conditioned Reflexes : An Investigation of the Physiological Activity of the Cerebral Cortex*. [http ://psychclassics.yorku.ca/Pavlov/index.htm](http://psychclassics.yorku.ca/Pavlov/index.htm)
- Peacocke, C. (1983). *Sense and Content*. Oxford : Clarendon Press.
- Peacocke, C. (1989). Perceptual content. In *Themes from Kaplan*, edited by P. Almog, and Wettstein. Oxford : Oxford University Press.
- Peacocke, C. (1992). *A Study of Concepts*. Cambridge, Mass. : MIT Press.
- Peacocke, C. (1998). Non-conceptual content defended (Comment on McDowell's 'Mind and World'). *Philosophy and Phenomenological Research* 58 :381-388.

Peacocke, C. (2001). Phenomenology and Non-conceptual Content. *Philosophy and Phenomenological Research* LXII (3) :609-615.

Prinz, J, J.(2002). *Furnishing the Mind*. Cambridge, Mass. : MIT Press.

Sedivy, S. (1996). Must conceptually informed perceptual experience involve non-conceptual content? *Canadian Journal of Philosophy* 26 :413-431.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, Mass. : MIT Press.

Ullman, S. (1996). *High-level Vision*. Cambridge, Mass. : MIT Press.

Bibliographie

Pierre Abélard. *Des intellections*. Vrin, 1994.

Martin Aigner. *Combinatorial Theory*. Springer, 1979.

Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change : Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2) :510–530, June 1985.

Scott Atran. *Cognitive foundations of natural history : towards an anthropology of science*. Cambridge University Press ; Édition de la Maison des sciences de l'homme, Paris, 1990.

Jeremy Avigad and Richard Zach. The epsilon calculus. URL <http://plato.stanford.edu/entries/epsilon-calculus/#3>.

Alfred J. Ayer. *Language, Truth and Logic*. Penguin Classics, 1936.

George Bealer. *Quality and Concept*. Oxford University Press, 1982.

Paul Benacerraf and Hilary Putnam, editors. *Philosophy of Mathematics, Selected readings*. Cambridge University Press, 1983 edition, 1964,.

Johan Van Benthem. *A Manual of Intensional Logic*. Noth-Holland Publishing Company, second edition, revised and expanded edition, 1988.

Johan Van Benthem. *Language in Action, Categories, Lambdas and Dynamic Logic*. Noth-Holland Publishing Company, 1991.

J.-P. Benzécri and collaborateurs. *L'analyse des données*, volume 1. La taxinomie. Dunod, 1976.

George Berkeley. *Principles of Human Knowledge (1710)*. Pinguin Classics, 1988.

- José Louis Bermúdez. *The Paradox of Self-Consciousness*. The MIT Press, 1998.
- José Louis Bermúdez. Nonconceptual content : From perceptual experience to subpersonal computational states. In York H. Gunther, editor, *Essays on Nonconceptual Content*. The MIT Press, 2003a.
- José Louis Bermúdez. Peacocke's argument against the autonomy of non-conceptual representational content (1994). In York H. Gunther, editor, *Essays on Nonconceptual Content*. The MIT Press, Cambridge, 2003b.
- Akeel Bilgrami. *Belief & Meaning*. Blackwell Publishing, 1992.
- Partick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, 2001.
- Simon Blackburn. *Ruling Passions, A Theory of Practical Reasoning*. Clarendon Press, Oxford, 1998.
- Erik Blaser, Zenon W. Pylyshyn, and Alex O. Holcombe. Tracking an object through feature space. *Nature*, 408 :196–199, 2000.
- Ned Block. Semantics, conceptual role. URL <http://www.nyu.edu/gsas/dept/philofaculty/block/papers/ConceptualRoleSemantics.html>.
- Paul Bloom. Possible individuals in language and cognition. *Current Directions in Psychological Science*, 50 :90–94, 1996.
- Paul Bloom. *Descartes's Baby : how the science of child development explains what makes us human*. Basic Books, 2004.
- Béla Bollobás. *Modern Graph Theory*. Springer, 1998.
- George Boolos. *Logic, Logic and Logic*. Harvard University Press, 1998.
- George S. Boolos, John P. Burgess, and Richard C. Jeffrey. *Computability and Logic*. Cambridge University Press, fourth edition, 2002.
- David Branddom-Mitchell and Frank Jackson. *Philosophy of Mind and Cognition*. Blackwell, 1996.
- Robert B. Brandom. *Making it Explicit : Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994.

- Robert B. Brandom. Précis of making it explicit. *Philosophy and Phenomenological Research*, LVII(1) :153–156, March 1997a.
- Robert B. Brandom. Replies. *Philosophy and Phenomenological Research*, LVII(1) :189–204, March 1997b.
- Robert B. Brandom. *Articulating Reasons, An Introduction to Inferentialism*. Harvard University Press, 2000.
- Bill Brewer. *Perception and Reason*. Clarendon Press, Oxford, 1999.
- Tyler Burge. Individualism and the mental. In P. French, T.E. Uehling, and H.K. Wettstein, editors, *Studies in Metaphysics*, volume IV of *Midwest Studies in Philosophy*. University of Minnesota Press, Minneapolis, 1979.
- John P. Burgess and Gideon Rosen. *A Subject With No Object, Strategies for Nominalistic Interpretation of Mathematics*. Oxford University Press, 1997.
- Peter J. Cameron. *Sets, Logic and Categories*. Springer, 1998.
- Susan Carey. *Conceptual Change in Childhood*. The MIT Press, 1987.
- Susan Carey. Bootstrapping & the origin of concepts. *Dedalus*, 113(1) : 58–69, January 2004.
- Susan Carey and Fei Xu. Infant’s knowledge of objects : beyond object files and object tracking? *Cognition*, 80 :179–213, 2001.
- Gregory Carlson. A unified analysis of the english bare plural. *Linguistics and Philosophy*, (1) :159–219, 1977.
- Rudolf Carnap. *Foundations of Logic and Mathematics*, volume I-II of *International Encyclopaedia of Unified Science*. The University of Chicago Press, 1939.
- Rudolf Carnap. *Meaning and Necessity, A Study in Semantics and Modal Logic*. The University of Chicago Press, second edition edition, 1947, 1956, 1988.
- Rudolf Carnap. *La construction logique du monde*. Mathesis. Vrin, 2002.

- Gail A. Carpenter and Stephen Grossberg. Integrating symbolic and neural processing in a self-organized architecture for pattern recognition and prediction. In V. Honavar and L. Uhr, editors, *Artificial intelligence and neural networks : Steps toward principled predictions*, chapter XVII, pages 387–421. San Diego Academic Press, 1994.
- Robyn Carston. *Thoughts and Utterances, The Pragmatics of Explicit Communication*. Blackwell Publishing, 2002.
- Roberto Casati. Common sense, philosophical and theoretical notions of an object : some methodological problems. In *Barcelona ESPP-SPP Joint Conference*, 2004.
- David J. Chalmers and Frank Jackson. Conceptual analysis and reductive explanation. *The Philosophical Review*, 110(3) :315–360, July 2001.
- Noam Chomsky. *Rules and Representations*. Columbia University Press, 1980.
- Alonzo Church. *Introduction to Mathematical Logic*. Princeton University Press, 1956.
- Nino B. Cocchiarella. *Logical Investigations of Predication Theory and the Problems of Universals*. Bibliopolis, Napoli, 1986.
- Annalisa Coliva. *I Concetti, Teorie ed esercizi*. Carocci, Roma, 2004.
- Antoine Cornuéjols and Laurent Miclet. *Apprentissage artificiel, Concepts et algorithmes*. Eyrolles, Paris, 2003.
- Tim Crane. The nonconceptual content of experience. In Tim Crane, editor, *The Content of Experience : Essays on Perception*. Cambridge University Press, 1992.
- Tim Crane. The waterfall illusion, 1988. In York H. Gunther, editor, *Essays on Nonconceptual Content*. The MIT Press, 2003.
- Gabrielle Crocco and Andreas Herzig. Les opérations de changement basées sur le test de ramsey. In Pierre Livet, editor, *Révision des croyances*. Lavoisier, 2002.

- Alan Cruse. *Meaning in Language, An Introduction to Semantics and Pragmatics*. Oxford University Press, 2000.
- G Csibra, G Gergely, S Biro, O Koos, and M Brockbank. Goal attribution without agency cues : the perception of 'pure reason' in infancy. *Cognition*, 72(3) :237–267, Oct 1999.
- A. Cussin. The connectionist construction of concepts. In M. A. Boden, editor, *The Philosophy of Artificial Intelligence*. Oxford University Press, 1990.
- Nigel J. Cutland. *Computability, An Introduction to Recursive Function Theory*. Cambridge University Press, 1980.
- Donald Davidson. On the very idea of a conceptual scheme. In *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford, 1984, 2001.
- Donald Davidson. *Enquête sur la vérité et l'interprétation*. Editions Jacqueline Chambon, 1993.
- Donald Davidson. Truth and meaning (1967). In Peter Ludlow, editor, *Readings in the Philosophy of Language*. MIT, 1997.
- Donald Davidson. *Subjective, Intersubjective, Objective*. Clarendon Press, 2001.
- Daniel C. Dennett. *La conscience expliquée*. Editions Odile Jacob, 1993.
- Daniel C. Dennett. *Brainstorms, Philosophical Essays on Mind and Psychology*. Pinguin Books, 1997.
- Eric Dietrich. Analogy and conceptual change, or you can't step into the same mind twice. In Eric Dietrich and Arthur B. Markman, editors, *Cognitive Dynamics : Conceptual and Representational Change in Humans and Machines*, chapter 10, pages 265–294. Lawrence Erlbaum Associates, 2000.
- Eric Dietrich and Arthur B. Markman. Cognitive dynamics : Computation and representation regained. In Eric Dietrich and Arthur B. Markman, editors, *Cognitive Dynamics : Conceptual and Representational Change in Humans and Machines*, chapter 1, pages 5–30. Lawrence Erlbaum Associates, 2000a.

- Eric Dietrich and Arthur B. Markman, editors. *Cognitive Dynamics, Conceptual and Representational Change in Humans and Machines*. Lawrence Erlbaum Associates, Mahwah, New Jersey, London, 2000b.
- Guillame d'Ockham. *Somme de logique*. Mauvezin, 1988.
- Jérôme Dokic. Situated mental representations. URL <http://jeannicod.ccsd.cnrs.fr/documents/disk0/00/00/03/06/index.html>. 2002.
- Jérôme Dokic. *Qu'est-ce que la perception ?* Vrin, 2004.
- Jérôme Dokic. *Les formes de l'indexicalité, Langage et pensée en contexte*, chapter Indexicalité et ineffabilité. Édition rue d'Ulm, 2005.
- Fred Dretske. *Seeing and knowing*. Routledge, 1969.
- Fred Dretske. *Knowledge and the Flow of Information*. MIT Press, 1981.
- Fred Dretske. *Explaining Behavior*. The MIT Press, 1988.
- Fred Dretske. *Naturalizing the Mind*. CNRS and MIT Press, the jean nicod lectures edition, 1995.
- Fred Dretske. *Perception, knowledge, and belief : selected essays*. Cambridge University Press, 2000.
- Fred Dretske. Sensation and perception (1981). In York H. Gunther, editor, *Essays on Nonconceptual Content*. The MIT Press, 2003.
- Jacques Dubucs. Belief and acceptance : A logical point of view. In Pascal Engel, editor, *Believing and Accepting*, pages 159–177. Kluwer Academic Publishers, 2000.
- Michael Dummett. *Frege, Philosophy of Language*. Harvard University Press, second edition edition, 1981.
- Michael Dummett. *The Seas of Language*. Clarendon Press, Oxford, 1993.
- Michael Dummett. What is a theory of meaning? (1974). In Peter Ludlow, editor, *Readings in the Philosophy of Language*. MIT, 1997.
- Pascal Engel. *Identité et Référence*. PENS, 1985.

- Pascal Engel, editor. *Believing and Accepting*, volume 83 of *Philosophical Studies Series*. Kluwer Academic Publishers, 2000.
- Pascal Engel. Le contenu de la perception est-il conceptuel? In *La Perception*. Jacques Bouveresse and J.J. Rosat, 2003.
- Angeles Erana, Benjamin Q. Sylvand, and Dario Taraborelli. Tracking agency. submitted.
- W. K. Estes. *Classification and Cognition*. Oxford University Press, 1994.
- Gareth Evans. *The Varieties of Reference*. Clarendon Press, Oxford, 1982.
- Gareth Evans. Understanding demonstratives (1981). In Peter Ludlow, editor, *Readings in the Philosophy of Language*. MIT, 1997.
- Paul Feyerabend. Comment être un bon empiriste. plaidoyer en faveur de la tolérance en matière épistémologique. In Pierre Jacob, editor, *De Vienne à Cambridge*. Tel Gallimard, 1980.
- Harty H. Field. Conventionalism and instrumentalism in semantics. *Noûs*, (9) :375–405, 1975.
- Harty H. Field. Logic, meaning, and conceptual role. *The Journal of Philosophy*, (7) :379–409, July 1977.
- Kit Fine. *The Limits of Abstraction*. Clarendon Press, 2002.
- Jerry A. Fodor. *The Language of Thought*. Harvard University Press, 1974.
- Jerry A. Fodor. *Concepts, Where Cognitive Sciences Went Wrong*. Clarendon Press, Oxford, 1998.
- Jerry A. Fodor, M.F. Garret, E.C.T. Wlaker, and C.H. Parkes. Against definitions. *Cognition*, (8) :263–367, 1980.
- Jerry A. Fodor and Ernest Lepore. *Holism*. Blackwell Publishing, 1992.
- Gottlob Frege. *Écrits logiques et philosophiques*. Seuil, 1971.
- Gottlob Frege. *Idéographie*. Vrin, 1999.
- Daniel Gallin. *Intensional and Higher-Order Modal Logic, With Applications to Montague Semantics*. Noth-Holland Publishing Company, 1975.

- Michel Galmiche. *Sémantique linguistique et logique, Un exemple : la théorie de R. Montague*. PUF, Paris, 1991.
- Bernhard Ganter and Guy W. Mineau, editors. *Conceptual Structures : Logical, Linguistic, and Computational Issues*, 2000. 8th International Conference on Conceptual Structures, ICCS 2000, Darmstadt, Germany, August 2000, Springer.
- Peter Gärdenfors. Belief revision and relevance. In *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume Two : Symposia and Invited Papers, pages 349–365, 1990.
- Peter Gärdenfors. La révision des croyances et le test de ramsey pour les conditionnels (1986). In Pierre Livet, editor, *Révision des croyances*, pages 33–43. Lavoisier, 2002.
- Peter Geach. *Mental Acts*. Thoemmes Press, 1992.
- G Gergely and G Csibra. Teleological reasoning in infancy : the infant's naive theory of rational action. a reply to premack and premack. *Cognition*, 63 (2) :227–233, May 1997.
- G Gergely, Z Nadasdy, G Csibra, and S Biro. Taking the intentional stance at 12 months of age. *Cognition*, 56(2) :165–193, Aug 1995.
- James J. Gibson. *The Senses Considered as Perceptual Systems*. Greenwood Press, Publishers, Westport, Connecticut, 1966.
- James J. Gibson. The theory of affordances. In R. Shaw and J. Bransford, editors, *Perceiving, Acting and Knowing*. Erlbaum, 1977.
- James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.
- Jean-Yves Girard. *Proof Theory and Logical Complexity*, volume 1. Bibliopolis, 1987.
- Kurt Gödel. On a hitherto unutilized extension of the finitary standpoint (1958). In Solomon Feferman, Jr John W. Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean Van Heijenoort, editors, *Collected Works*, volume II, pages 240–251. Oxford University Press, 1990.

- Nelson Goodman. *Faits, Fictions et Prédications*. Les éditions de minuit, 1984.
- Nelson Goodman and Catherine Z. Elgin. *Reconceptions en philosophie*. PUF, 1994.
- A. Gopnik and A. N. Meltzoff. *words, Thoughts, and Theories*. The MIT Press, 1996.
- A. D. Gordon. *Classification, Methods for the exploratory analysis of multivariate data*. Chapman and Hall, London, 1981.
- Mark Greenberg and Gilbert Harman. Conceptual role semantics. URL www.princeton.edu/~harman/Papers/CRS.pdf.
- York H. Gunther. Content, illusion, partition. *Philosophical Studies*, 102 : 185–2002, 2001.
- York H. Gunther, editor. *Essays on Nonconceptual Content*. The MIT Press, Cambridge, 2003.
- Ian Hacking. *Le plus pur nominalisme, "vleu" et ses usages*. Édition de l'éclat, Combas, 1993.
- Bob Hale. *Abstract Objects*. Basil Blackwell, 1987.
- James Hampton. A demonstration of intransitivity in nautral categories. *Cognition*, (12) :151–164, 1982.
- James Hampton. Overextension of conjunctive concepts : Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology : Learning, Memory and Cognition*, (14) :12–32, 1988.
- James Hampton. Typicality and compositionality : The logic of combing vague concepts. URL http://www.cognition.ens.fr/nac2004/nac_abstracts.php#4.Typical. June 20 2004.
- Gilbert Harman. *Change in View, Principles of Reasoning*. The MIT Press, 1986.

- Gilbert Harman. (nonsolipsistic) conceptual role semantics. In Ernest Lepore, editor, *New Directions in Semantics*, pages 55–81. London : Academic Press, 1987.
- Gilbert Harman. *Reasoning, Meaning and Mind*. Oxford University Press, 1999.
- Jane Heal. *Fact and Meaning, Quine and Wittgenstein on Philosophy of Language*. Basil Blackwell, 1989.
- F. Heider and M. Simmel. An experimental study of apparent behavior. *American Journal of Psychology*, 57 :243–249, 1944.
- J. Heil. Perceptual experience. In B. McLaughlin, editor, *Dretske and his Critics*. Basil Blackwell, 1991.
- David Hilbert and Paul Bernays. *Fondements des mathématiques*, volume 1. L'Harmattan, 2001a.
- David Hilbert and Paul Bernays. *Fondements des mathématiques*, volume 2. L'Harmattan, 2001b.
- Paul Horwich. What it is like to be a deflationary theory of meaning. pages 133–154. Ridgeview Publishing Company, 1994.
- Paul Horwich. The composition of meanings. *The Philosophical Review*, 106 (4) :503–532, October 1997.
- G.E. Huges and M.J. Cresswell. *A new introduction to modal logic*. Routledge, 1996.
- David Hume. *Enquête sur l'entendement humain (1758)*. GF-Flammarion, 1983.
- S. Hurley. *Consciousness in Action*. The MIT Press, 1998.
- Pierre Jacob. *L'intentionnalité, Problèmes de philosophie de l'esprit*. Odile Jacob, 2004.
- Pierre Jacob. Recognitional concepts and conceptual combination. In Markus Werning, Edouard Machery, and Gerhard Schurz, editors, *The Composition of Meaning and Content.*, volume I : Foundational Issues. Ontos Verlag, 2005.

- Pierre Jacob and Marc Jeannerod. *Ways of Seeing, The Scope and Limits of Visual Cognition*. Oxford University Press, 2003.
- Richard Jeffrey. *Probability and the Art of Judgment*. Cambridge University Press, 1992.
- Richard Jeffrey. Transformation of belief. In Pascal Engel, editor, *Believing and Accepting*, pages 145–157. Kluwer Academic Publishers, 2000.
- Jean-François Jodouin. *Les réseaux neuromimétiques*. Hermes, 1994.
- Philip Nicholas Johnson-Laird. *Mental models : toward a cognitive science of language, inference and consciousness*. Cambridge University Press, 1983.
- Philip Nicholas Johnson-Laird and Ruth M. J. Byrne. *Deduction*. Hove, 1991.
- D. Kahneman and Ann Treisman. The reviewing of object files : Object-specific integration of information. *Cognitive Psychology*, 24 :175–219, 1992.
- Hans Kamp. A theory of truth and semantic representation. In Paul Portner and Barbara H. Partee, editors, *Formal Semantics, The Essential Readings*. Blackwell Publishing, 2002.
- Makoto Kanazawa, Christopher Piñón, and Henriëtte de Swart, editors. *Quantifiers, Deduction, and Context*. CSLI, 1995.
- Emmanuel Kant. *Critique de la Raison Pure*. PUF, 2004.
- Jerrold J. Katz. Semantics and conceptual change. *The Philosophical Review*, 88(3) :327–265, July 1979.
- Jerrold J. Katz. *The Methaphysics of Meaning*. The MIT Press, 1990.
- Jerrold J. Katz and Paul M. Postal. Realism vs. conceptualism in linguistics. *Linguistics and Philosophy*, (14) :515–554, 1991.
- S. Kelly. Demonstrative concepts and experience. *Philosophical Review*, 110 (3) :397–420, July 2001.

- Ildiko Kiraly, Bianca Jovanovic, Wolfgang Prinz, Gisa Aschersleben, and Gyorgy Gergely. The early origins of goal attribution in infancy. *Conscious Cogn*, 12(4) :752–769, Dec 2003.
- Saul Kripke. *La Logique des Noms Propres (Naming and Necessity)*. Les éditions de Minuit, 1982 edition, 1972, 1980.
- Jean-Louis Krivine. *Lambda-calcul, types et modèles*. Masson, 1990.
- Jean-Louis Krivine. *Théorie des ensembles*. Cassini, Paris, 1998.
- V.A. Kuhlmeier, P. Bloom, and K. Wynn. Do 5-month-old infants see humans as material objects? *Cognition*, 94(1) :95–103, 2004a.
- V.A. Kuhlmeier, K. Wynn, and P. Bloom. People vs. objects : a reply to rakison and cicchino. *Cognition*, 94(1) :109–112, 2004b.
- Thomas Kuhn. *The Structure of Scientific Revolutions*. The University of Chicago Press, 3d edition, 1962, 1970, 1996.
- Thomas Kuhn. La tension essentielle : tradition et innovation dans la recherche scientifique. In Pierre Jacob, editor, *De Vienne à Cambridge*. Tel Gallimard, 1980.
- René Lalement. *Logique, réduction, résolution*. Études et recherches en informatique. Masson, Paris, 1990.
- Mark Lance. Quantification, substitution, and conceptual content. *Noûs*, 30(4) :481–507, December 1996.
- Daniel Laurier. *Introduction à la philosophie du langage*. Mardaga, 1980.
- Daniel Laurier and François Lepage, editors. *Essais sur le langage et l'intensionnalité*. Bellarmin, Vrin, 1992.
- Guillaume Lecointre and Hervé LeGuyader. *Classification phylogénétique du vivant*. Belin, Paris, 2001.
- Gottfried Wilhelm Leibniz. *Dissertatio de arte combinatoria*. Francfort, 1690.
- Gottfried Wilhelm Leibniz. *Nouveaux essais sur l'entendement humain*. Flammarion, Paris, 1990.

- François Lepage and Serge Lapierre. *Logique partielle et savoir, Essai de philosophie formelle*. Bellarmin Vrin, 2000.
- A.M. Leslie, Fei Xu, Patrice D. Tremoulet, and Brian J. Scholl. Indexing and the object concept : developing what and where systems. *Trends in Cognitive Sciences*, 2 :10–16, 1998.
- Isaac Levi. Subjunctives, dispositions and chances. *Synthese*, (34) :432–455, 1977.
- Harold D. Levin. *Categorical Grammar and the Logical Form of Quantification*. Bibliopolis, Napoli, 1982.
- Clarence Irving Lewis. *An Analysis of Knowledge and Valuation*. The Open Court Publishing Company, La Salle, Illinois, 1946.
- Clarence Irving Lewis. *Mind and the World-Order, Outline of a Theory of Knowledge*. Dover Publication, 1956.
- Clarence Irving Lewis. A paradox of nominalism, june 16 1953. In John D. Goheen and John L. Mothershead Jr, editors, *Collected Papers*, page 331. Stanford University Press, 1970a.
- Clarence Irving Lewis. Realism or phenomenalism? In John D. Gohenn and John L. Mothershead Jr, editors, *Collected Papers*, pages 335–347. Stanford University Press, 1970b.
- Clarence Irving Lewis and Cooper Harold Langford. *Symbolic Logic*. Dover Publication, 1932.
- David Lewis. *Convention*. Blackwell, 2002 edition, 1969.
- David Lewis. *Philosophical Papers*, chapter General Semantics, pages 189–233. Oxford University Press, 1983a.
- David Lewis. *Philosophical Papers*, volume 1 and 2. Oxford University Press, 1983b.
- David Lewis. *On the Plurality of Worlds*. Blackwell, 1986.
- David Lewis. *Papers in Philosophical Logic*. Cambridge University Press, 1998.

- David Lewis. *Papers in Metaphysics and Epistemology*. Cambridge University Press, 1999.
- Pierre Livet, editor. *Révision des croyances*. Lavoisier, Paris, 2002.
- John Locke. *An Essay Concerning Human Understanding (1689)*. Penguin Classics, 1997.
- Saunders Mac Lane. *Homology*. Classics in Mathematics. Springer, 1994 edition, 1963, 1975.
- Diego Marconi. *Lexical Competence*. The MIT Press, Cambridge, 1997.
- Ruth Barcan Marcus. A backward look at quine's animadversion on modalities (1990). In *Modalities, Philosophical Essays*. Oxford University Press, 1993a.
- Ruth Barcan Marcus. Classes, collections, assortments, and individuals (1963). In *Modalities, Philosophical Essays*. Oxford University Press, 1993b.
- Ruth Barcan Marcus. Does the principle of substitutivity rest on a mistake? (1975). In *Modalities, Philosophical Essays*. Oxford University Press, 1993c.
- Ruth Barcan Marcus. Essential attribution (1971). In *Modalities, Philosophical Essays*. Oxford University Press, 1993d.
- Ruth Barcan Marcus. Essentialism in modal logic (1967). In *Modalities, Philosophical Essays*. Oxford University Press, 1993e.
- Ruth Barcan Marcus. Iterated deontic modalities (1966). In *Modalities, Philosophical Essays*. Oxford University Press, 1993f.
- Ruth Barcan Marcus. Modalities and intensional languages (1961). In *Modalities, Philosophical Essays*. Oxford University Press, 1993g.
- Ruth Barcan Marcus. *Modalities, Philosophical Essays*. Oxford University Press, 1993h.
- Ruth Barcan Marcus. Moral dilemmas and consistency (1980). In *Modalities, Philosophical Essays*. Oxford University Press, 1993i.

- Ruth Barcan Marcus. Nominalism and the substitutional quantifier (1978). In *Modalities, Philosophical Essays*. Oxford University Press, 1993j.
- Ruth Barcan Marcus. On some post-1920s views of Russell on particularity, identity, and individuation (1986). In *Modalities, Philosophical Essays*. Oxford University Press, 1993k.
- Ruth Barcan Marcus. Possibilia and possible worlds (1986). In *Modalities, Philosophical Essays*. Oxford University Press, 1993l.
- Ruth Barcan Marcus. Quantification and ontology (1972). In *Modalities, Philosophical Essays*. Oxford University Press, 1993m.
- Ruth Barcan Marcus. Rationality and believing the impossible (1983). In *Modalities, Philosophical Essays*. Oxford University Press, 1993n.
- Ruth Barcan Marcus. Some revisionary proposals about belief and believing (1990). In *Modalities, Philosophical Essays*. Oxford University Press, 1993o.
- Ruth Barcan Marcus. Spinoza and the ontological proof. In *Modalities, Philosophical Essays*. Oxford University Press, 1993p.
- Eric Margolis and Stephen Laurence, editors. *Concepts, Core Readings*. The MIT Press, Cambridge, 1999.
- M.G.F. Martin. Perceptual content. In S. Guttenplan, editor, *A Companion to the Philosophy of Mind*. Basil Blackwell, 1994.
- Per Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, Napoli, 1984.
- Per Martin-Löf. On the meanings of the logical constants and the justifications of the logical laws. *Nordic Journal of Philosophical Logic*, 1(1) : 11–60, 1996.
- John McDowell. Brandom on representation and inference. *Philosophy and Phenomenological Research*, LVII(1) :157–162, March 1997.
- John McDowell. *Mind and World*. Harvard University Press, 2002.
- D. Medin and M. Schaffer. Context theory of classification learning. *Psychological Review*, 85(3) :207–238, 1978.

- C. Meini. La frontière entre percepts et concepts - la théorie de dretske à l'épreuve des illusions d'optique. *Incognito*, 6(27-38), 1996.
- Ruth Garrett Millikan. *language, Thought, and other Biological Categories : New Foundations for Realism*. MIT, 1984.
- Jean-Maurice Monnoyer, editor. *La structure du monde, objects, propriétés, états de chose, Renouveau de la métaphysique dans l'école australienne de philosophie*. Vrin, 2004.
- Richard Montague. *English as a Formal Language*, pages 189–224. Edizioni di Comunità, Milan, 1970a.
- Richard Montague. Universal grammar. *Theoria*, (36) :373–398, 1970b.
- Richard Montague. The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsik, and P. Suppes, editors, *Approaches to Natural Language : Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, pages 221–242, Dordrecht, 1973. Reidel.
- Gregory Murphy. *The Big Book of Concepts*. The MIT Press, 2002.
- John F. Nash. Equilibrium points in n-person games. In *Proceedings of National Academy of Science*, volume 36, pages 48–49, 1950.
- Bernhard Nebel. *Reasoning and Revision in Hybrid Representation System*. Springer-Verlag, Heidelberg, New York, 1995.
- Alva Noë. *Action in Perception*. The MIT Press, 2004.
- Elisabeth Pacherie. Levels of perceptual content. *Philosophical Studies*, (100) :237–254, 2000.
- J. Paris and L. Harrington. A mathematical incompleteness in peano arithmetic. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 1133–1142. North-Holland, 1977.
- I.P. Pavlov. Conditioned reflexes : An investigation of the physiological activity of the cerebral cortex, 1927. URL <http://psychclassics.yorku.ca/Pavlov/index.htm>.
- Christopher Peacocke. *Sense and Content*. Oxford University Press, 1983.

- Christopher Peacocke. *Thoughts, An Essay on Content*. Blackwell Publishing, 1986.
- Christopher Peacocke. *A Study of Concepts*. The MIT Press, 1992.
- Christopher Peacocke. Does perception have a nonconceptual content ? URL <http://humanities.ucsc.edu/NEH/peacocke2/pdf>. 1999.
- Christopher Peacocke. Nonconceptual content : Kinds, rationales, and relations (1994). In *Essays on Nonconceptual Content*, chapter 16, pages 309–322. The MIT Press, 2003a.
- Christopher Peacocke. Scenarios, concepts, and perception. In York H. Gunther, editor, *Essays on Nonconceptual Content*. The MIT Press, 2003b.
- Christopher Peacocke. *The Realm of Reason*. Oxford, 2004.
- Charles Sanders Peirce. *Le raisonnement et la logique des choses*. Les Éditions du Cerf, Paris, 1995.
- Aline Péliissier and Alain Tête, editors. *Sciences cognitives, Textes fondateurs (1943-1950), Wiener, Rosenblueth, Bigelow, McCulloch, Pitts, von Neumann, Hebb, Weaver, Shannon, Turing*. PUF, 1995.
- John Perry. *Knowledge, Possibility, and Consciousness*. The 1999 Jean Nicod Lectures. The MIT Press, Cambridge, 2001.
- Paul Portner and Barbara H. Partee, editors. *Formal Semantics, The Essential Readings*. Blackwell Publishing, 2002.
- Daniel Povinelli, James E. Reaux, Laura A. Theall, and Steve Giambrone. *Folk physics for apes : the chimpanzee's theory of how the world works*. Oxford University Press, 2003.
- Dag Prawitz. *Natural Deduction, A Proof-Theoretical Study*. Almqvist & Wiksell, Stockholm, 1965.
- D. Premack. The infant's theory of self-propelled objects. *Cognition*, 36(1) : 1–16, 1990.
- Graham Priest. *An Introduction to Non-Classical Logic*. Cambridge University Press, 2001.

- Jesse Prinz. *Furnishing the Mind : Concept and Their Perceptual Basis*. The MIT Press, 2002.
- Arthur Prior. *Objets de pensée*. Vrin, 2002.
- Arthur Prior. *Papers on Time and Tense*. Oxford University Press, 2003.
- Hilary Putnam. Meaning and reference. *The Journal of Philosophy*, 70(19) : 699–711, November 8 1973.
- Hilary Putnam. The meaning of "meaning". In K. Gunderson, editor, *Language, Mind and Knowledge : Minnesota Studies in the Philosophy of Science*, volume 70. University of Minnesota Press, Minneapolis, 1975.
- Hilary Putnam. *Raison, Vérité et Histoire*. Les éditions de Minuit, 1981, 1984.
- Hilary Putnam. *Représentation et réalité*. NRF essais Gallimard, 1990.
- Hilary Putnam. *Philosophie de la logique*. L'éclat, Combas, 1996.
- Zenon Pylyshyn. *Seeing and Visualizing*. The MIT Press, 2003.
- Zenon Pylyshyn. Jean nicod lectures 2004. 2004a.
- Zenon Pylyshyn. Visual indexes, objects, and nonconceptual reference. 2004b.
- Zenon W. Pylyshyn. Situating vision in the world. *Trends in Cognitive Sciences*, 4 :197–207, 2000.
- Zenon W. Pylyshyn. Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80 :127–158, 2001.
- Zenon W. Pylyshyn. Notes for the oleron summer workshop on objects. June 2004c.
- Willard Van Quine. Toward a calculus of concepts. *The Journal of Symbolic Logic*, 1(1), 1936.
- Willard Van Quine. *From a Logical Point of View*. Harvard University Press, 1953.

- Willard Van Quine. Introduction to "on the building blocks of mathematical logic" by Moses Schönfinkel. In Jean van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879-1931*, pages 355–357. Harvard University Press, 1976.
- D.H. Rakison and J.B. Cicchino. Is an infant a people person? *Cognition*, 94(1) :105–107, 2004.
- Frank Plumpton Ramsey. *Philosophical Papers*. Cambridge University Press, 1990, 1994.
- Frank Plumpton Ramsey. *Notes on Philosophy, Probability and Mathematics*. Bibliopolis, Napoli, 1991.
- Aarne Ranta. *Type-Theoretical Grammar*. Clarendon Press, 1994.
- François Recanati. *Direct Reference*. Blackwell Publishing, 1993.
- Ronald A. Rensink. The dynamic representation of scenes. *Visual Cognition*, 2000.
- Irvin Rock. *The Logic of Perception*. The MIT Press, 1983.
- Richard Rorty. What do you do when they call you a 'relativist'? *Philosophy and Phenomenological Research*, LVII(1) :173–177, March 1997.
- Eleanor Rosch. Cognitive representation of semantic categories. *Journal of Experimental Psychology*, (104) :573–605, 1975.
- Gideon Rosen. Who makes the rules around here? *Philosophy and Phenomenological Research*, LVII(1) :163–171, March 1997.
- Jay F. Rosenberg. Brandom's *Making It Explicit* : A first encounter. *Philosophy and Phenomenological Research*, LVII(1) :179–187, March 1997.
- Hans Rott. *Change, Choice and Inference, A Study of Belief Revision and Nonmonotonic Reasoning*. Clarendon Press, Oxford, 2001.
- Bertrand Russell. *Principles of Mathematics*. Routledge, (1903),(1937), 2002.
- Bertrand Russell. *Introduction à la philosophie mathématique*. Payot, 1991 edition, 1921.

- Bertrand Russell. *Ecrits de logique philosophique*. PUF, 1989.
- Daniel L. Schacter and Elain Scarry, editors. *Memory, Brain, and Belief*. Harvard University Press, 2000.
- Stephen Schiffer. *The Things We Mean*. Oxford University Press, 2003.
- Brian J. Scholl and Zenon W. Pylyshyn. Tracking multiple items through occlusion : Clues to visual objecthood. *Cognitive Psychology*, pages 259–290, 1999.
- Brian J. Scholl, Zenon W. Pylyshyn, and Jacob Feldman. What is a visual object : Evidence from target-merging in multiple-object tracking. *Cognition*, 80(1-2) :159–177, 2001.
- Brian J. Scholl, Zenon W. Pylyshyn, and Steve L. Franconeri. The relationship between property-encoding and object based attention : Evidence from multiple object tracking, 1999. URL <http://rucss.rutgers.edu/ftp/pub/papers/scholl-pylyshyn-franconeri.pdf>.
- Brian J. Scholl and P. Tremoulet. Perceptual causality and animacy. *Trends Cogn Sci*, 4(8) :299–309, 2000.
- Moses Schönfinkel. On the building blocks of mathematical logic (1924). In Jean van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879-1931*, pages 357–366. Harvard University Press, 1976.
- John R. Searle. Proper names. *Mind*, 67(266) :166–173, april 1958.
- S. Sedivy. Must conceptually informed perceptual experience involve non-conceptual content? *Canadian Journal of Philosophy*, 26 :413–431, 1996.
- Wilfrid Sellars. *Science, Perception and Reality*. Ridgeview Publishing Company, Atascadero, California, 1991.
- Wilfrid Sellars. *Empirisme et philosophie de l'esprit*. L'éclat, Combas, 1992.
- Stewart Shapiro. *Philosophy of Mathematics, Structure and Ontology*. Oxford University Press, 1997.
- Joseph R. Shoenfield. *Mathematical Logic*. Association for Symbolic Logic, A K Peters Ltd, Natick, Massachusetts, 1967.

- Peter Simons. *Parts, A Study in Ontology*. Oxford University Press, 1987.
- John Skorupski. Reasons : the basic concepts. April 11th 2004.
- John Skorupski. Propositions about reasons. *Journal of Philosophy*, to appear. URL http://www.st-andrews.ac.uk/academic/philosophy/STAFF/Propositions_about_reasons.pdf.
- Edward E. Smith and Douglas L. Medin. *Categories and Concepts*. Harvard University Press, 1981.
- Elizabeth Spelke. Origins of visual knowledge. In D.N. Osherson, S.M. Kosslyn, and J.M. Hollerbach, editors, *Visual Cognition and Action*. MIT Press, Cambridge, Mass., 1990.
- Elizabeth Spelke. Object perception. In Alvin I. Goldman, editor, *Readings in Philosophy and Cognitive Science*. MIT Press, Cambridge, Mass., 1993.
- Elizabeth Spelke. Initial knowledge : Six suggestions. *Cognition*, 50 :431–445, 1994.
- Dan Sperber, editor. *Metarepresentations, A Multidisciplinary Perspective*. Oxford University Press, 2000.
- Dan Sperber and Deirdre Wilson. *Relevance, Communication and Cognition*. Harvard University Press, 1986.
- Robert C. Stalnaker. *Inquiry*. MIT Press, 1984.
- Robert C. Stalnaker. *Context and Content*. Oxford University Press, 1999.
- P. F. Strawson. *Individuals, An Essay in Descriptive Metaphysics*. Routledge, 1959.
- Göran Sundholm. Inference versus consequence. In *LOGICA Yearbook*, chapter Inference versus Consequence. Czech Acad.Sc, Prague, 1998a.
- Göran Sundholm. Semantic values for natural deduction derivation. January 1998b.
- Göran Sundholm. Maccoll on judgement and inference. *Nordic Journal of Philosophical Logic*, pages 119–132, 1999.

- Benjamin Sylvand. La querelle des idées abstraites entre locke et leibniz et ses prolongements dans le débat contemporain. Master's thesis, Paris 4, 2000.
- Benjamin Sylvand. Le problème des abstraits mathématiques chez frege, husserl et dans la philosophie de l'arithmétique contemporaine. Master's thesis, Paris 4, 2001.
- Benjamin Sylvand. La logique des taxinomies. Master's thesis, Paris 4 and IRCAM, 2002.
- Benjamin Sylvand. Reference and inference. In *Conference on Robert Brandom's Philosophy*, 2005.
- Benjamin Sylvand and Nivedita Gangopadhyay. The notion of the non-conceptual in an explanation of illusion. submitted.
- Gaisai Takeuti. On a generalized logic calculus. *Japanese Journal of Mathematics*, XXIII :39–96, 1953.
- Alfred Tarski. *Logic, Semantics, Meta-Mathematics*. Hackett, 1956.
- Alfred Tarski, Andrzej Mostowski, and Raphael M. Robinson. *Undecidable Theories*. Noth-Holland Publishing Company, Amsterdam, 1953.
- Richmond H. Thomason, editor. *Formal Philosophy : Selected Papers of Richard Montague*. Yale University Press, 1974.
- William T. Trotter. *Combinatorics and Partially Ordered Sets, Dimension Theory*. Johns Hoptkins Series in the Mathematical Sciences, Baltimore and London, 1992.
- Alan Turing. Computing machinery and intelligence. *Mind*, (59) :433–460, 1950.
- A. Tversky and I. Gati. Similarity, separability, and the triangle inequality. *Psychological Review*, (89) :123–154, 1982.
- A. Tversky and D. Kahneman. Judgments of and by representativeness. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment Under Uncertainty : Heuristics and Biases*. Cambridge University Press, 1982.

- Amos Tversky. Features of similarity. *Psychological Review*, (84) :327–352, 1977.
- Michael Tye. *Ten Problems of Consciousness*. The MIT Press, 1996.
- Michael Tye. *Consciousness, Color, and Content*. MIT Press, 2000.
- S. Ullman. *Hight-Level Vision*. The MIT Press, 1996.
- K. Van Marle and Brian J. Scholl. Attentive tracking of objects versus substances. *Psychological Science*, 14(5) :498–504, September 2003.
- Georg Henrik von Wright. *The Logical Problem of Induction*. The MacMillan Company, New York, 1957.
- David Wiggins. *Sameness and substance renewed*. Cambridge University Press, 2001.
- Timothy Williamson. *Vagueness*. Routledge, London and New York, 1994.
- Timothy Williamson. *Knowldge and its Limits*. Oxford University Press, 2000.
- Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell, 1958.
- Andrew Woodfiled. Conceptions. *Mind*, 100(4) :547–572, October 1991.
- Andrew Woodfiled. Do your concepts develop? In C. Hookway and D. Peterson, editors, *Philosophy and Cognitive Science*, pages 41–67. Cambridge University Press, 1993.
- Fei Xu. Object individuation and object identity in infancy : The role of spatiotemporal information, object property information, and language. *Acta Psychologica*, 102 :113–136, 1999.
- Fei Xu and Susan Carey. Infants’ metaphysics : The case of numerical identity. *Cognitive Psychology*, 30 :111–153, 1996.
- Lofti A. Zadeh. Fuzzy sets. *Inf. Control*, (8) :338–3365, 1965.
- Edward N. Zalta. A classically-based theory of impossible worlds. *Notre Dame Journal of Formal Logic*, 38(4) :640–661, Fall 1997.
- Denis Zwirn. Principe de miller et révision des croyances probabilistes. In Pierre Livet, editor, *Révision des croyances*. Lavoisier, 2002.

Index

- abduction, 96
- abstraction, 86, 87, 95, 99
 - principe d'abstraction, 88
 - idées abstraites, 159
 - opérateur d'abstraction, 177
 - principe d'abstraction, 91
 - processus d'abstraction, 84
- abstrait, 55, 56, 63, 73, 77, 84, 88, 92, 96, 109, 173, 176
 - caractère abstrait, 146
 - entité abstraite, 145
 - objet abstrait, 161
 - terme abstrait, 177
- acquisition, 59, 64, 83, 103, 108, 116
- action, 53, 65
- AGM, 209
- air de famille, 89
- analyticit , 150
- appartenance, 73
- application, 60, 97, 164, 168, 210
- apprentissage, 59, 87, 150, 152, 156, 165
 - apprentissage artificiel, 63
 - apprentissage automatique, 61
- assignation, 38
- atomique, 54
- atomisme, 47, 52, 56, 66, 103
 - atomisme conceptuel, 111, 117
 - atomisme logique, 108
- attention, 42
- attribution de concepts, 218
- auto-r f rentialit , 48
- axiomatique, 174
- Berkeley, 98
- c blage, 53
 - pr c blage, 53
- calcul des pr dicats, 176
- canonique, 67
- capacit  conceptuelle, 146, 152, 166
- caract risation, 173
- caract re, 55, 69, 85, 87, 96, 157, 174
- cat gorisation, 48, 63
- cat rsienanisme, 56
- cause, 51
- changement
 - changement de concept, 39, 209
 - changement de conception, 217
 - changement de croyance, 209
- classification, 69, 72, 110
- cloture logique, 209
- cognitif, 67

- cohérence, 170, 174
combinaison, 54, 63
combinatoire, 54, 109, 156, 157,
159–161
communication, 64, 65, 106, 107,
115
compétence, 153
 domaine de compétence, 151,
152
comparaison, 84
composition, 54, 108
compositionnalité, 56, 60, 62, 81,
93, 97, 105, 109, 110, 155,
156, 159–161
compréhension, 218
concept
 concept empirique, 116
 concept complexe, 109, 156,
157
 concept composé, 108
 concept primitif, 108
 concept primitif, 109
 nouveau concept, 156, 159
conception, 67, 112, 211, 216
conceptualisme, 72, 75
concevable, 160
concordance extentionnelle, 157
condition d'application, 146
connaissance, 42, 75
connecteurs, 64
connotation, 174
construction de concept, 161, 162
constructivisme, 174
construire, 113
contenu
 contenu étroit, 116
 contenu cognitif, 83, 85, 91,
97, 103, 145
 contenu conceptuel, 35, 36, 39,
87
 contenu large, 116
 contenu non-conceptuel, 35,
36, 39, 87, 91, 100, 107,
116
 contenu objectif, 93, 104, 145
 contenu sémantique, 162
 contenu scénario, 116
 non-conceptuel, 141, 146
contexte, 92, 164, 170
contraction, 209, 212
 contraction de concept, 213
 contraction de croyance, 213
contradiction, 174
contrainte de généralité, 44, 146
correction, 38, 39, 52
croyance, 41

décomposition, 156
déduction naturelle, 168
déférence, 152
définition, 69, 82, 147, 162, 174
dénotation, 44, 158, 160, 170
dérivation
 dérivation cognitive, 67
dériver, 113
désignation, 55
Davidson, 218
descripteur, 63
description, 154, 155
 descripteur, 156
discrimination, 42, 43, 62
disposition, 85

- division du travail linguistique, 65, 67
 domaine, 47, 78, 112, 114, 118, 146, 156, 159, 176
 domaine d'application, 153
 donnée sensorielle, 92
 Dretske, 42
 effectivité, 175
 effet de typicité, 73, 83, 88–90, 103
 empirique, 177
 empirisme, 61, 83, 87, 98
 empirisme conceptuel, 105
 ensemble, 48, 61, 88, 162, 173
 sous-ensemble flou, 88
 sous-ensemble flou, 93
 théorie des ensembles, 91, 93
 épistémique, 41, 52, 75, 87, 115, 165
 épistémique fort, 116
 principe épistémique faible, 81
 principe épistémique fort, 81
 épistémique
 caractère épistémique faible, 41
 caractère épistémique fort, 41
 caractère non-épistémique faible, 43
 caractère non-épistémique fort, 43
 épistémicité, 63
 essentialisme, 79
 état du monde, 160
 Evans, 44
 exemplaire, 82
 expérience, 77, 82, 116, 141, 143, 144, 146, 150, 153
 expansion, 209
 expansion de concept, 210
 expansion de croyance, 209
 expert, 68, 80, 115
 extension, 48, 86, 87, 91, 105, 110, 111, 115, 116, 145, 146, 157–159, 161–164, 175–177, 210
 co-extension, 73, 144, 157, 162
 extension idéale, 177
 extensionnel, 144
 externalisme, 41
 Feyerabend, 217
 fiablisme, 101
 fonction, voir fonctionnalisme
 fonction caractéristique, 45, 55, 61–63, 73, 175, 210
 fonction de composition, 58
 fonction de correction, 38
 fonctionnalisme
 fonction, 179
 forme logique du concept, 150, 156, 157, 160, 161
 forme propositionnelle, 162, 166
 généralisation, 86, 92, 96
 généralité, 145, 146
 générativité, 107, 108
 granularité, 85
 hiérarchie, 72, 91, 110
 holisme, 40, 41, 47, 50, 60, 78, 80, 90, 104, 111, 114, 117, 153, 162, 164, 170
 homologie, 50, 157
 homonymie, 117

- Hume, 98
- idéalisation, 177
- Idée, 70, 98, 100
- εἶδος, 70
- identification, 43, 62, 63, 77, 98, 145, 146
- identifier, 146
- identité, 77, 145
- identité de Levi, 215
- idolect, 115
- illusion, 104
 - illusion de constance de taille, 35–37
- image mentale, 83, 99
- inclusion, 156
- incommensurabilité, 81, 218
- indentification, 173
- individuation, 46, 47, 53, 98, 102, 103, 145, 146, 173
- inférence, 63, 78, 87, 99, 156, 160
 - capacité inférentielle, 142, 149, 165, 166
 - réseau inférentiel, 142, 145, 150, 170, 176
 - règle d'inférence, 169
 - règles d'inférence, 159
 - schéma d'inférence, 177
 - schéma inférentiel, 174
 - structure inférentielle, 148, 151, 153, 159, 161–163, 165, 175
 - trame inférentielle, 150, 154–159, 161, 165, 175, 177, 179
- inférence
 - structure inférentielle, 113
 - relation inférentielle, 112
 - trame inférentielle, 114
- inférences, 64
- inné, 59
 - innéisme, 153
- innéisme, 60, 61, 66, 106, 107, 109
- innéisme
 - anti-inisme, 103
- instance, 82, 146, 150, 169, 170
- intelligence artificielle, 61, 63
- intension, 41, 45, 98, 111, 115, 116, 145, 146, 158, 159, 161, 164, 174, 176, 177
 - co-intension, 102, 144
 - intensionnel, 92, 144
- intention, 106
- intentionnalité, 106
- internalisme, 41
- interprétation, 218
- intersubjectif, 68
- Kuhn, 217
- label, 61
- λ calcul, 172
- lexical, 145, 149, 157
- Locke, 98
- logique, 87
- logique combinatoire, 55, 172
- mémoire, 83
- modèle, 160
- mode de présentation, 43, 53, 62, 93, 146, 150
- monde actuel, 160, 176
- monde possible, 58, 73, 111, 160, 176

- nativisme, 46, 60, 106, 109
 anti-nativisme, 103
- nombre de Gödel, 50
- nominalisme, 50, 75, 85
- non-conceptuel, 51, 53, 60, 62–64,
 75, 92, 113, 115, 143, 150,
 153
- non-contradiction, 159, 160, 175
- norme, 38, 66, 113, 114, 116
- objectif, 68
- objet
 objet mathématique, 160, 177
- omniscience, 39, 41, 72, 108
- ontologie
 statut ontologique, 160
- opérateur ϵ , 175–177
- opérateur λ , 176, 177
- opérateur γ , 175
- opérateur d'abstraction, 45
- opération, 179
- paradigmes, 217
- paradoxe de Russell, 48
- particulier, 77
- percept, 39, 60, 61, 63, 87, 104, 116
- perception, 60, 61
- pertinence, 114, 164
- poids, 95
- point d'équilibre, 88, 89
- portée, 146
- possession, 46, 103, 110, 111, 113
- postuler, 67, 113
- poto-objet, 60
- pré-conceptuel, 60
- prédicat, 73, 173, 175, 176
- préjugé, 78
- présentation, 146, 150
- pragmatique, 85
- pragmatisme, 56
- preuve, 40
- primitif, 59, 61
- principe contrastif, 49, 62, 84
- principe d'identité, 62
- principe de dérivation canonique
 postulée, 218
- principe de référence canonique
 postulée, 218
- proposition, 52
- propriété, 44, 45, 62, 63, 69, 72, 82,
 145, 146, 148, 154, 157–
 159, 173, 176, 177, 210
 propriété cachée, 79
 propriété saillante, 39, 53, 60–
 63
- proto-catégorisation, 87
- proto-concept, 60, 62
- proto-jugement, 53, 62, 113, 115,
 116
- proto-jugements, 87
- proto-proposition, 52, 116
- prototypes, 86, 88
- proxytype, 61, 62, 102–104
- publicité, 64, 105–107, 109, 115,
 149, 165
 caractère public, 148, 151, 152
- Putnam, 218
- qualie, 116
- quantificateur, 176
- réalisation, 175
- réduction, 96

- référence, 41, 44, 48, 55, 65, 105, 116, 146, 158, 170
cadre référentiel, 157
co-référence, 93, 102
référence canonique, 61
référence causale, 105, 112
référence cognitive, 61, 67
référence objective, 112
référence postulée, 58
téléologie de la référence, 112
- réflexe, 53
- révision, 209, 215
révision de concept, 154, 155, 216
révision de conception, 216
révision de croyance, 215
- rôle inférentiel, 58
- règle d'élimination, 166
- règle d'introduction, 166
- règle de Bayes, 155
- règle de déduction, 143
- raisonnement, 168
- Ramsey, 51
- rationalité, 209
- rationnel, 218
- recognitionnel, 177
- reconnaissance, 62, 77, 98
- relations, 64
- relativisme, 40, 104
- représentation, 37, 38, 52, 63, 87, 112, 115, 146, 170, 175
- représentationnel, 37, 41
- ressemblance, 73
- routine, 51, 53
- sémantique, 51, 66
sémantique des rôles conceptuels, 179, 216
- scénario, 116
- scène, 60
- schéma, 172
- schème conceptuel, 218
- sens, 44, 45, 67
- sens commun, 68, 79
- similitude, 49, 84, 88, 90, 92
- solipsisme, 107, 218
- stéréotype, 89
- substitution, 147, 157
substitution extensionnelle, 156
- substrat, 173
- synonymie, 117
- téléosémantique, 195
- taxinomie, 72
- terme général, 77
- terme lexical, 65
- théorie, 82
- théorie classique, 82, 91, 92, 98, 105
- théorie classique des concepts, 112, 162, 174
- théorie des exemplaires, 49, 91, 92, 96, 99, 105
- théorie des prototypes, 49, 56, 86, 103, 105
- théorie empirique, 49, 57, 116
- théorie naïve, 47, 68, 77, 91, 92, 112
- théorie scientifique, 47, 75
- théorie-théorie, 75, 82, 105
- tiers exclu, 69, 72
- tiers-exclu, 183

- traduction, 218
- trait, 62, 88
- trait caractéristique, 39, 49
- trame conceptuelle, 112
- transitivité, 72, 88, 110

- universaux, 98, 100
- universel, 77
- usage, 60, 85, 210
 - mésusage, 66
- usage cognitif, 116

- vérificationisme, 39

Concept et changement de concept
Concept, contenu et inférence,
Bases pour une approche dynamique du concept

Benjamin Sylvand

Le concept est l'entité mentale qui permet à un agent cognitif qui la possède de penser son environnement. Le concept se caractérise par deux dimensions : d'une part son application (les objets auxquels il réfère) et son usage d'autre part (les relations qu'il entretient avec les autres états mentaux). Les théories contemporaines du concept ne parviennent pas à rendre compte de ces deux aspects. Cela vient de l'assimilation de la théorie des concepts avec une théorie de la classification et d'une distinction floue entre le conceptuel et le nonconceptuel. Le concept est considéré ici comme étant représentationnel alors que le nonconceptuel ne l'est pas. Représentationnel signifie ici que le concept peut être changé ou corrigé alors que le nonconceptuel (le percept) ne l'est pas. La théorie du concept défendue ici est normative dans la mesure où des principes sont énoncés pour caractériser le concept. L'agent possède un concept s'il est capable de postuler une assignation à celui-ci. Et ce à la fois en première personne (contenu cognitive) et en troisième personne (contenu canonique). Cela suppose une théorie de l'esprit et un accès au contenu du concept (aspect épistémique). L'agent doit également pouvoir rendre compte de l'usage qu'il fait du concept, et ce non seulement pour lui (dérivation cognitive) mais aussi en troisième personne (dérivation canonique). Le holisme qui découle de cette conception implique que le concept apparaît toujours dans une « conception », c'est-à-dire comme inférentiellement relié à d'autres concepts. Cette théorie du concept permet de rendre compte du changement et de l'évolution d'un concept plutôt que son remplacement systématique.

Concept is the mental entity that allows the cognitive agent who possesses it to think about his environment. It is characterized by two dimensions : its application (objects it refers to) and its use (relationships it bears with other mental states). The contemporary theories of concept fail to explain both these features of concept, due to the confusion made between a theory of concept and a theory of classification and of a fuzzy distinction between conceptual and nonconceptual. The main difference between conceptual and nonconceptual is that the former is representational whereas the latter is not. Being a representation means being corrigible. The theory of concept proposed here is normative. Principles are offered for characterizing the concept. A cognitive agent possesses a concept if he is able to assign a content to it at first person (cognitive content) and at third person (canonical content). This entails a theory of mind. Moreover, the agent must be able to explain his use of the concept, for himself (cognitive derivation) and for other minds (canonical derivation). The holism implied by this theory enables us to explain the change and the evolution of concept which are different from concept replacement.
