UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA CONCEPTUALISATION COMME INTERFACE ENTRE RÉSEAUX DE NEURONES ET STRUCTURE SOCIALE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAITRISE EN PHILOSOPHIE

PAR
FÉLIX BRUNETTA

MARS 2014

AVANT-PROPOS

Ce mémoire est l'aboutissement d'un long processus de domestication de mon corps juvénile. Merci à mon grave accident de voiture de m'avoir forcé à plonger dans mes lectures, à mes enfants de m'avoir montré la valeur du temps en me le faisant perdre à jouer aux figurines. Merci à mes parents pour ce fond de culpabilité qui m'a maintenu à l'école. Merci à ma compagne, Marie-Flavie, qui m'inspire vers le haut. Merci à mon directeur, Pierre Poirier, qui est demeuré à mes côtés virtuellement, tel mon daïmon, m'apostrophant lorsque je quittais le fleuve de la rationalité. Merci également à Serge Robert pour ses lectures et ses commentaires. Finalement, merci à Jean Guy Meunier de m'avoir fourni la clé qui motive toute cette entreprise, à savoir le concept de concept.

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES TABLEAUX

«Les individus, et même des peuples entiers, ne pensent guère que, pendant qu'ils poursuivent leurs intentions privées, chacun selon ses goûts, et souvent contre les autres individus, ils suivent comme un fil directeur, sans s'en apercevoir, l'intention de la nature, qui leur est inconnue, et qui, même s'ils en avaient connaissance, leur importerait cependant peu».

Kant 1784

RÉSUMÉ

Depuis une quarantaine d'années, plusieurs nouvelles théories sur les concepts proviennent de la psychologie à en croire la littérature[1]. On y distingue souvent trois théories des concepts qui s'y font compétition: la théorie des exemples, des prototypes et celle des théories. Selon nous, ces théories ne décrivent pas totalement les concepts. D'une part, ces théories psychologiques réduisent la nature des concepts aux traitements que des individus en font: acquisition, reconnaissance, identification, etc. Jamais on y explique leur création, leur origine. Or, un individu seul, sans supervision, ne peut construire de concept. Partant, comment conjuguer le fait que d'une part l'apprentissage est supervisé et orienté par des concepts déjà existants et d'autre part que ces concepts sont décrits et expliqués comme la production d'individus? Comment expliquer que les agents isolément ne possèdent pas les capacités cognitives suffisantes pour construire les concepts que l'on connait (turbo réacteur par exemple), mais que ces mêmes capacités sont les seuls outils que l'on a pour expliquer la construction de concepts? La solution est peut-être à chercher à un niveau social: la production d'un concept est peut-être le fruit de l'interaction entre plusieurs individus d'une population. Certains modèles de populations de réseaux de neurones semblent conforter cette piste de solution.

MOTS-CLÉS : concept, origine des concepts, réseaux de neurones, réseaux sociaux, internalisme.

[1] Machery(09), Margolis(99), Murphy(02), Weiskopf(09)

INTRODUCTION

Le cerveau est un être essentiellement désintéressé par sa propre existence. Plus sérieusement, il constitue un organe dont la fonction est d'être à propos d'autre chose que lui-même. Il reçoit des signaux des différents récepteurs sensoriels en périphérie, il les intègre, les emmagasine et il produit des réponses. Ainsi, il est sensible aux choses qui l'entourent, mais il ne peut pas, physiologiquement, se sentir lui-même. À l'image des yeux qui ne se verront jamais directement, un cerveau ne peut pas faire l'expérience directe de son activité physique.

Le travail du cerveau n'est pas seulement à propos des signaux pour eux-mêmes, mais dans la mesure où ces signaux sont causés par d'autres événements dans le monde, ces derniers en véhiculent la trace. Les signaux réfèrent à des choses autres que leur nature électrique. L'exemple des caméras pour aveugle illustre bien ceci. Ces dernières captent d'abord la lumière réfléchie par des objets sur une matrice à pixels et ensuite retransmettent la configuration d'activité de ces pixels sur une autre matrice située sur la langue d'un individu. Cette dernière matrice produit dans la bouche des signaux électriques correspondant aux patrons d'activité sur la caméra qui, à leur tour, aboutissent dans le cerveau par voie nerveuse. Au départ, les aveugles qui apprennent à utiliser ces appareils ne perçoivent qu'un chatouillement chaotique sur la langue; ils ne sentent que les signaux eux-mêmes. Mais, après quelques séances d'entrainement, ils apprennent à reconnaitre à quoi réfèrent les différents patrons d'activité électrique. Ils apprennent à *voir*.

Ainsi, que ce soit le patron d'activité sur le capteur de la caméra, celui émulé sur les cellules de la langue ou encore celui produit par une assemblée de neurones dans le cerveau, certaines choses servent à en représenter d'autres. Maintenant, on constate qu'il existe de multiples façons de représenter le monde, c'est-à-dire qu'il y a différentes choses qui peuvent *se tenir à la place* d'une autre et ce, de différentes manières. Par exemple, il y a une différence entre la complexité de la configuration à un moment x (aussi court que puisse être un moment) de l'excitation des capteurs d'une rétine oculaire qui représente une pomme particulière et la configuration de l'activité d'une population de neurones qui implémente le concept POMME. La première représentation possède une relation plus directe avec sa référence. Elle est une impression physique sur la rétine causée par la lumière réfléchie depuis la pomme dans le monde. En ce sens elle est éphémère. À chaque fois qu'un nouveau photon frappe la rétine, le patron d'activité des cônes et bâtonnets de cette dernière forme théoriquement une nouvelle représentation. Un œil peut ainsi former des milliards de représentations particulières d'une même chose. Au contraire, la représentation POMME est plus loin de la sensibilité directe (elle émerge d'une séquence plus nombreuse d'opérations cognitives) dans la mesure où elle ne réfère pas à une pomme particulière, mais à toutes les pommes à la fois. En d'autres termes, elle est plus abstraite et ainsi elle offre une plus grande stabilité par rapport au monde en constant changement. Le concept POMME possède une durée de vie supérieure aux différentes représentations d'une pomme particulière, il possède même une existence plus longue que les pommes physiques.

Maintenant, le caractère abstrait des représentations conceptuelles leur confère des propriétés de généralisation. En effet, si l'opération de représentation au niveau des sens est à propos d'éléments très localisés, ces représentations ne sont pas intégrées de manière globale dans un système représentationnel, mais demeure plutôt fonction de l'environnement toujours en changement. Pourtant ce dont nous sommes conscients semble davantage organisé et intégré qu'un chaos pixellisé et mouvant.

À l'autre extrême, une représentation au niveau supérieur, voire le plus abstrait possible, correspond à des formes générales de la pensée, à des invariants non situés directement dans le monde. Prenons l'exemple d'une tasse bleue qui aurait son anse orientée vers la gauche et une tasse rouge qui aurait son anse vers la droite. Ces deux choses correspondent au même contenu représentationnel supérieur[2]: l'invariant TASSE. Ici, encore une fois, les contenus de ce niveau ne ressemblent pas plus à ce qui a lieu dans notre perception consciente. Nous ne voyons pas l'invariant TASSE isolément lorsque nous faisons l'expérience de deux tasses singulières et différentes. C'est plutôt à un niveau plus intermédiaire que les représentations atteignent un équilibre entre le *localisé* et l'*abstrait* pour former ce que l'on perçoit consciemment : des perceptions qui sont particulières et localisées dans le monde tout en étant organisées globalement dans une structure de pensées. Or, cette faculté d'intégrer du particulier à l'aide d'invariants plus abstraits est une opération de généralisation dans la mesure où certaines représentations plus abstraites possèdent des propriétés partagées et attribuables à une multitude de représentations singulières concrètes. Dans ce sens, il est possible de subsumer ces dernières sous les premières.

C'est précisément cette synthèse de représentations singulières et diverses sous une représentation générale que la tradition depuis Aristote appelle un concept. *Le concept résulte de ce qu'il y a de commun entre plusieurs* choses [3] ou plusieurs représentations singulières.

Soit, depuis une quarantaine d'années les innovations théoriques sur les concepts proviennent de la psychologie et des sciences cognitives à en croire la littérature[4]. On y distingue souvent trois théories des concepts qui s'y font compétition: la

[2] J.Prinz, *A Neurofunctional Theory of Consciousness,* 2003, University of North Carolina, Chapel Hill

[3] Schmid A-F. *Concept,* in Dictionnaire de Philosophie et des sciences. Ed, PUF. Paris. 2003.

[4] Machery(09), Margolis(99), Murphy(02), Weiskopf(09)

théorie des *exemples*, des *prototypes* et celle des *théories*. Ces théories psychologiques définissent généralement la nature des concepts en fonction des traitements que des individus en font: acquisition, construction, reconnaissance, identification, etc. Dans ce sens, les concepts sont des entités internes et propres à des individus. En effet, nous l'avons vu, ils sont compris comme des entités informationnelles qui émergent en aval d'une série d'opérations de représentation interne à un sujet.

Un problème survient toutefois avec le fait de définir les concepts comme des choses produites essentiellement à l'intérieur d'individus pris isolément. Si on est d'accord avec l'idée qu'un enfant ne peut s'éduquer lui-même en confrontant ses mécanismes représentationnels et son environnement, mais qu'il est préférable de lui enseigner des versions officielles du français, des mathématiques, de l'histoire, de l'informatique, etc.; alors d'où viennent les concepts appris dans ces matières? En d'autres termes, comment expliquer le fait que l'apprentissage soit supervisé et orienté par des concepts déjà existants, et que ces concepts soient paradoxalement produits par des individus? Qu'est-ce qui est arrivé en premier, le concept ou l'individu? Comment expliquer que les agents cognitifs humains pris isolément ne possèdent pas les capacités cognitives suffisantes pour construire les concepts que l'on connait, TURBO RÉACTEUR par exemple, mais que ces agents humains sont les seuls appareils cognitifs connus (mis à part les artéfacts robotiques conçus pour imiter l'humain) susceptibles d'expliquer la construction de concept? Bref, la question de ce mémoire est : comment expliquer que des individus humains ou des robots les imitant peuvent créer des concepts, des structures sémantiques qui pourront jouer le rôle des représentations conceptuelles? Notre thèse est que bien que la conceptualisation soit opérée par des mécanismes de catégorisation propre aux individus, il est erroné d'attribuer la source des contenus des concepts à des agents cognitifs seuls ou pris isolément. La solution est peut-être à chercher à un niveau social. La production d'un concept est peut-être le fruit de l'interaction entre

plusieurs individus d'une population : c'est-à-dire, qu'un concept est possiblement construit à partir du commerce de représentations devenues publiques et externes. Partant, les concepts auraient eux-mêmes la particularité de transcender les individus et d'être collectifs.

Dans ce mémoire, nous exposons cette piste de solution au problème de l'origine des concepts en quatre chapitres. Primo, nous définissons ce qu'est un concept avec une attention particulière aux nouvelles théories en sciences cognitives et leur rapport avec la théorie classique. Deuzio, après s'être engagé sur une définition du concept de concept, nous cherchons à expliquer la création de ces derniers depuis leur origine, et ce, en tenant compte de la nécessité d'invoquer une architecture cognitive physique pour ainsi répondre à nos prétentions matérialistes et naturalistes. Les modèles de réseaux de neurones classiques sont retenus ici pour leur capacité à produire des représentations suivant un travail en partie *buttom up* et pour les analogies potentielles avec l'humain. Tertio, nous montrons l'impossibilité pour un agent cognitif seul ou pris isolément de répondre au défi de la création de concepts. Dans ce chapitre nous évoquons longuement les raisons historiques et sociales qui ont mené les sciences cognitives vers ce cul-de-sac internaliste et individualiste. Finalement, nous proposons à l'instar de Hutchins et Hezelhurst[5] une piste de solution selon laquelle une population de réseaux de neurones semblent être capable de créer des représentations généralisantes plus près des véritables concepts que ne le peut un individu seul.

---

[5] Hutchins et Hazlehurst, *Learning in the Cultural Process*, Department of Cognitive ScienceUniversity of California, 1990

CHAPITRE I

CONCEPT DE CONCEPT

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

Lorsque je pense à une chose tel qu'un chien, je possède une *image mentale*, mais aucun chien n'est réellement dans ma tête[6]. Comment la connaissance intelligible de la matière est-elle possible?

Les réponses traditionnelles véhiculent l'idée que les choses sont re-présentées dans l'esprit (mind) en tant que forme. Un état ou une entité à l'intérieur de l'agent cognitif sert de substitut formel, parfois immatériel, à un objet matériel extérieur. Les représentations, les idées, l'intentionnalité, les informations, les concepts sont des exemples de substituts formels.

De plus, certaines représentations semblent être plus que le substitut d'un objet particulier dans le monde et portent plutôt sur une multitude de choses plus ou moins semblables. Par exemple, je peux me représenter ce qu'est une fraise en général et l'appliquer à plusieurs occurrences de fraises dans l'environnement. Ici FRAISE est

---

[6] Meunier J-G, *Trios types de représentations*. Publication du LANCI. UQAM. 2002. www.unites.uqam.ca/lanci/

une représentation générale regroupant plusieurs caractéristiques essentielles que partagent plusieurs représentations ou exemplaires singuliers de fraises.

Cette synthèse de représentations singulières et diverses sous une représentation générale est ce que la tradition depuis Aristote appelle un concept. *Le concept résulte de ce qu'il y a de commun entre plusieurs choses* [7] ou plusieurs représentations singulières.

1.1- Concept en philosophie

1.1.1-Théorie Classique

Dans le même sens, plusieurs théoriciens tels que Margolis[8] (1999) résument l'histoire du concept de concept depuis l'Antiquité à l'aide de ce que l'on nomme la *Théorie Classique.* Cette théorie générale pose les concepts comme des structures définitionnelles qui encodent les conditions nécessaires et suffisantes pour leurs propres applications. De cette façon, les concepts sont des structures représentationnelles complexes, possédant des conditions d'application qui sont en fait de plus simples représentations (des caractéristiques essentielles). Par exemple, John Locke définissait le concept de soleil comme un agrégat de plusieurs représentations simples : brillant, chaud, rond, qui bouge de façon régulière, qui est à une certaine distance de nous, etc.[9]

Ultimement, du moins pour les empiristes, ces représentations singulières qui forment les conditions nécessaires et suffisantes à l'application des concepts se décomposent à leur tour en unités sensorielles et/ou perceptuelles (*features*). Si l'on

---

[7] Schmid A-F. *Concept*, in Dictionnaire de Philosophie et des sciences. Ed, PUF. Paris. 2003.

[8] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence 1999

[9] *Ibid*

inclut cette thèse empiriste à la *Théorie Classique*, acquérir un concept revient à assembler un complexe de *features* de telle façon que quelque chose tombe sous ce concept si et seulement s'il satisfait ces *features.* De cette manière, tous les concepts sont, à la fin, déterminés par un ensemble de représentations sensorielles.

Cette description est utile pour expliquer la capacité de classifier et de généraliser qu'offrent les concepts. En effet, réaliser un complexe de conditions pour juger de l'éligibilité d'une chose à un ensemble est une bonne explication de l'opération de classification. Elle explique aussi la généralisation: on peut partir avec son concept de fraise, par exemple, et cueillir une multitude d'objets qui possèdent les *features* essentiels; cueillir des objets qui tombent sous ce concept.

En résumé, on peut définir la notion classique de concept comme une structure définitionnelle qui encode un ensemble de conditions nécessaires et suffisantes à leur application, idéalement (pour les empiristes) des conditions en termes sensoriels et/ou perceptuels. Ainsi la conceptualisation est une opération de classification des choses communes; et finalement ces classes sont généralisantes dans la mesure où elles concernent une multitude d'occurrences.

1.1.2- Autres critères traditionnels en philosophie

Ces dernières caractéristiques décrivent le concept de manière endogène. Toutefois, dans l'histoire de la philosophie, plusieurs philosophes ne partagent pas l'idée strictement empiriste que les concepts sont ultimement composés de *features* exprimant des propriétés sensorielles, certains considèrent plutôt que l'identité du concept de concept est redevable des relations qu'il entretient avec d'autres aspects de l'esprit, les aspects exogènes. Rapidement, voici deux critères ou aspects exogènes qu'un modèle devrait réaliser, selon un autre pan de la tradition, pour conceptualiser.

1) D'abord, la conceptualisation n'est pas qu'une réaction à l'environnement, elle est une opération de représentation qui n'est pas appliquée aux objets externes directement, ou répondant à des stimuli, mais à d'autres représentations. Par exemple, elle traite des exemplaires, des représentations simples, des *features*, etc. C'est une opération de synthèse à partir d'autres représentations. 2) De plus, les concepts participent à d'autres opérations. Mentionnons le jugement, le raisonnement, le contrôle de la perception et le langage. Ici, les concepts ne sont pas que des classes isolées, ils participent en réseau à la cognition supérieure en général. En d'autres termes, l'esprit n'est pas un réservoir de classes comme Google, il réalise toujours les concepts dans des opérations plus complexes.

Tableau 1.1: caractéristiques traditionnelles des concepts en philosophie

Critères endogènes	Critères exogènes
opération de classification à l'aide d'une définition	Représentation supérieure (hors de l'environnement)
ces classes sont généralisantes et intégrantes	les concepts participent à d'autres opérations

1.2-Critique de la Théorie Classique

Durant la deuxième moitié du 20^{ième} siècle, la Théorie Classique de concept a été sévèrement critiquée. Ce n'est pas le but du travail de couvrir 60 ans de débats, mais un bref survol des principales critiques suffira pour expliquer l'impopularité du concept de concept durant ces dernières années.

Primo, en 1953, Wittgenstein montre, avec le concept de jeu, la difficulté de saisir des définitions nécessaires et suffisantes pour les concepts[10]. En occurrence, les critères importants de la définition de jeu sont tous confrontés à des contre-exemples : activité compétitive (contre–exemple : les jeux de cartes tel que le solitaire), implique de gagner ou perdre (contre-exemple : jouer à se lancer la balle), etc. Si l'on ne peut trouver de définition généralisante, comment circonscrire un concept alors?

En fait, on peut faire remonter ce problème aux *Dialogues* de Platon, alors que Socrate remettait toujours en doute les concepts de ses interlocuteurs et montrait la difficulté de définir un concept. La nouveauté avec Wittgenstein[11] n'est pas seulement le fait que nous soyons incapables de formuler des définitions, c'est que nous n'en avons pas besoin. Nous utilisons avec succès des concepts sans pouvoir en donner de définition. Par exemple, même s'il est *impossible* de définir le «jeu», nous jouons et reconnaissons facilement les jeux. Par conséquent, pour comprendre un concept, il faut regarder comment il fonctionne dans un contexte spécifique et non chercher des caractéristiques nécessaires et suffisantes. Maintenant, comprendre un concept dans un contexte n'est pas toujours un processus conscient. Pour le Wittgenstein des *Investigations,* nous catégorisons intuitivement en *remarquant* les ressemblances entre les choses; les ressemblances familiales. Par exemple, en regardant deux personnes et leurs différents traits physiques (nez, yeux, taille, etc.), nous pouvons intuitivement nous rendre compte s'ils sont des frères ou non. Ce processus est généralisé à toute catégorisation. Bref, le sens des termes ou des concepts est dérivé de leur utilisation par des humains.

Deuzio, les *ressemblances familiales* ont trouvé des applications dans différents domaines hors de la philosophie. Entre autres, elles ont influencé les travaux de la

---

[10] Margolis, *Concepts core reading,* Edited by Eric Margolis and Stephen Laurence 1999

[11] Wittgenstein, *Investigation philosophiques*. 1953

psychologue Eleanor Rosch[12]. Celle-ci observe que même lorsqu'une définition semble garantir l'application théorique du concept à tout coup, dans les faits, les gens ne semblent pas toujours utiliser les concepts en comparant des définitions et des choses. Les gens semblent habiles pour juger de la typicalité d'une chose par rapport à une catégorie. Rosch a su montrer que l'on produit facilement des classements de typicalité concernant plusieurs concepts[13]. Plus un item possède des caractéristiques que devraient posséder les autres de la même catégorie, plus on lui attribue de points sur une échelle (de 1 à 7 par exemple). Ainsi un moineau est un oiseau plus typique qu'un kiwi ou un pingouin. De plus, la typicalité affecte la vitesse de réponse, en plus d'affecter le taux d'erreur de jugement catégoriel. Plus un item est typique, plus on le classe rapidement et moins on se trompe. Le problème pour la *Théorie Classique* est qu'elle ne peut pas expliquer ces phénomènes et donc qu'elle n'a pas su les prédire. Car si les concepts possèdent des critères nécessaires et suffisants à leur application, les items admis devraient tous être de bons exemples également.

Tertio, avec peu d'exemples solides de définitions nécessaires et suffisantes et suite aux critiques empiriques, la *Théorie Classique* repose beaucoup sur sa capacité d'expliquer des phénomènes sémantiques tels que les inférences analytiques. L'exemple le plus connu d'inférence analytique est *Smith est un célibataire, alors Smith est un homme.* Ici, l'inférence semble garantie par la définition de célibataire. En effet, les critères homme, pas marié, etc., sont inclus dans la définition avant toute application du concept et garantissent leur utilisation.

---

[12] Rosch E. and Mervis, C. *Family resemblances: studies in the internal structure of categories*, Cognitive Psychology **7**, 1975

Rosch, E., *Wittgenstein and categorization research in cognitive psychology*, in M. Chapman & R. Dixon (Eds.), *Meaning and the growth of understanding. Wittgenstein's significance for developmental Psychology*, Hillsdale, NJ.: Erlbaum. 1987.

[13] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence. 1999.

On n'apprend rien sur le monde avec une inférence analytique, ce sont des connaissances a priori ou des tautologies comme l'a précisé le Cercle de Vienne. C'est dans ce sens que selon la *Théorie Classique,* chaque concept est circonscrit par une définition analytique.

Or, dans son célèbre article «Deux Dogmes de l'Empirisme», W. Quine critique avec beaucoup d'impact l'existence d'une différence entre des énoncés analytiques et synthétiques. Quine montre qu'aucun énoncé ne possède d'ensemble isolable de conditions à son adhésion qui pourrait être établi a priori, sans possible révision. Ces conditions sont plutôt inhérentes à un système holistique d'énoncés soumis à des mouvements, des changements. En d'autres termes, même les concepts qui semblent immunisés contre toute critique peuvent apparaitre comme rejetables suite à des développements théoriques. Par exemple, le concept de «ligne droite» semblait contenir la caractéristique d'«être le chemin le plus court entre deux points». Cependant les géométries post-euclidiennes ont fait sauter ce concept, qui semblait pourtant solide[14].

Bref, Quine a montré qu'il n'y avait pas de distinction soutenable entre l'analytique et le synthétique et donc les concepts ne peuvent pas être définissables de la façon que la *Théorie Classique* le requiert.

1.3 – Nouvelle vie pour le concept de concept en psychologie

Suite aux nombreuses critiques du modèle classique de concept, ce dernier a été occulté par plusieurs philosophes lors de la deuxième moitié du 20ième siècle. Les solutions aux problèmes sont plutôt venues de la psychologie, de la linguistique, des

[14] Murphy, *The big book of concept*, Massachusetts Institute of Technology. 2002

autres sciences de la cognition et des philosophes qui s'intéressent à ces domaines[15]. Ces différentes disciplines ont porté leur regard plus empiriste sur la nature des concepts et ont mis l'accent sur la manière qu'a un individu de conceptualiser plutôt que sur les définitions en soi. Ici, un concept est a priori une entité mentale et non une entité abstraite[16]. Dans ce contexte, il faut maintenant tenir compte tant des propriétés représentationnelles (du contenu), que des propriétés fonctionnelles (comment les concepts sont utilisés dans les processus cognitifs). C'est-à-dire que les propriétés attribuées aux concepts doivent s'accommoder à la façon qu'ont les individus d'utiliser et d'acquérir des concepts lors de tâches expérimentales.

Plus précisément, en psychologie, on s'entend pour affirmer que les concepts sont des structures représentationnelles mises en réserve dans la mémoire à long terme et sont utilisés par différents processus de la cognition supérieure tels que la catégorisation, l'induction, la déduction, etc.[17] Ce sont en quelques sortes les *ensembles* de bases de la connaissance (bodies of knowledge) communes à toutes ces opérations de la cognition supérieure. Dans le même sens, selon Machery, l'on soutient en psychologie que les gens catégorisent et raisonnent de la façon qu'ils le font puisque ces processus utilisent des concepts qui possèdent toutes les mêmes propriétés. Par conséquent, définir les propriétés générales que partagent tous les concepts mémorisés est un préalable à une théorisation de la cognition supérieure.

Ce nouvel engouement pour le concept de concept a produit trois principales familles de théories psychologiques: les théories des *exemples*, des *prototypes* ou des *théories*[18].

---

[15] La philosophie de l'esprit s'est aussi intéressée au concept de concept, mais il n'en sera pas question dans ce texte étant donné l'espace restreint.

[16] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence 1999

[17] Edouard Machery, *Concepts Are Not a Natural Kind*,

[18] Selon Piccinini : Voir Hampton [1993] pour les prototypes, Nosofsky [1988] pour les exemples, et

1.3.1-Théorie des exemples

La théorie des exemples a d'abord été proposée par Medin et Schaffer (1978)[19]. Selon ces derniers, un concept se résume à un ensemble de représentations particulières mémorisées qui partagent des caractéristiques assez proches pour qu'on puisse déceler leur similitude. Par exemple, le concept de chien correspond à toutes les représentations particulières de chiens qu'un individu mémorise. Lorsqu'on voit un nouvel objet (ici un chien) dans notre environnement, on le compare à des exemples de notre mémoire suivant des règles de calcul de similitude et on l'associe aux exemples les plus semblables. Non pas à une représentation regroupant des caractéristiques essentielles.

C'est précisément l'originalité de cette théorie qu'il n'y ait pas de représentation générale qui synthétiserait l'ensemble des cas particuliers. En fait, il n'y a pas de concept tel que le conçoit la tradition. Néanmoins, plusieurs problèmes qu'avait rencontrés la théorie classique sont résolus. Principalement, la nécessité d'établir une liste de caractéristiques nécessaires et suffisantes propres à une catégorie ne tient plus. Bien au contraire, les ensembles d'exemples se créent et se transforment selon les choses perçues dans le temps. De plus, cette continuité dans le processus de conceptualisation laisse place à un certain flou concernant l'éligibilité à un concept. Par exemple, la théorie prédit des *exemples* à cheval sur plusieurs ensembles; un *exemple* pourrait partager un même nombre de caractéristiques avec deux ou plusieurs concepts. Dans le même sens, la théorie des *exemples* a une explication

---

Gopnik and Meltzoff [1997] pour les théories

[19] Medin, *Concepts and conceptual structure*, 1989, in Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence 1999

pour le phénomène de typicalité soulevé par Rosch[20]. Plus un *exemple* ressemble à un grand nombre de membres d'un ensemble, plus il est typique. Une représentation particulière de labrador est un *exemple* de chien qui partage des caractéristiques avec beaucoup d'autres représentations de cet ensemble et peu avec d'autres ensembles tels que celui des marmottes ou des rats. Une représentation de chihuahua, par contre, ressemble à moins d'*exemples* de chiens que le labrador et peut ressembler à des *exemples* de rats[21].

1.3.2 - Théorie des prototypes

Bien que l'on attribue l'origine des théories décrivant explicitement les *prototypes* à celle de Rosch et Mervis 1978, il en existe plusieurs à ce jour. Dans tous les cas, un *prototype* n'est pas une représentation d'une chose en particulier comme l'est un *exemple*. Il est une représentation abstraite dont la structure encode une synthèse statistique des propriétés que ses membres tendent à posséder, plus particulièrement, par la plus grande partie des membres de son ensemble. Ceci dit, les propriétés synthétisées dans un prototype ne sont pas nécessaires et suffisantes. Un membre d'un ensemble/concept n'est pas obligé d'instancier toutes les propriétés du prototype, mais quelques-unes.

En effet, dans la mesure où les *prototypes* sont des structures conceptuelles mises en réserve dans la mémoire à long terme, un sujet applique un concept à une chose perçue en calculant (inconsciemment) si cette chose perçue possède un nombre de propriétés typiques d'un concept, selon sa mémoire, à un degré suffisant. Par exemple, si Obama voit un oiseau, il en forme une représentation particulière qui contient des propriétés propres aux oiseaux. Obama compare alors les propriétés de

---

[20] Voir la section «critique de la théorie classique»

[21] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence 1999

cette représentation particulière avec les propriétés représentées par différents *prototypes* de sa mémoire, incluant celui des oiseaux. La similitude entre l'oiseau vu par Obama et chaque *prototype* est calculée. Et puisque la représentation particulière partage davantage de propriétés avec le prototype «oiseau» que tous les autres, Obama conclut que ce qu'il voit est un oiseau.

Cette possibilité de calculer la similitude règle le problème du phénomène de typicalité que rencontrait la théorie classique. Plus un membre possède de propriétés représentées par le prototype, plus il est typique. De plus, il règle les problèmes du manque de définition et d'analycité puisqu'il n'exige pas de définition nécessaire et suffisante.

1.3.3-Théorie des théories

Selon la théorie des *théories*, l'acquisition et l'utilisation des concepts ne reposent pas sur un calcul de la similarité comme pour les *prototypes* et les *exemples*. Murphy et Medin (1985)[22], par exemple, supposent que les concepts fonctionnent comme des théories scientifiques et, par conséquent, que les opérations cognitives qui leurs sont associées fonctionnent comme des théories scientifiques. Ceci implique que les concepts ne se résument pas à une liste ou une synthèse de propriétés. Au contraire, l'identité d'un concept dépend de ses relations avec les autres concepts. Par exemple, le concept de mouffette, au lieu de se résumer à une liste de propriétés perceptibles comme: noire, ligne blanche sur le dos, pue, etc., comprend plutôt des explications sur des propriétés ou liens cachés comme: est un mammifère, a du sang, un code génétique, n'existe pas sur la lune, ne miaule pas, etc. C'est de cette façon que l'on ne se laisse pas berner par nos *prototypes* ou nos

[22] Murphy, *The big book of concept,* Massachusetts Institute of Technology 2002

exemples comme la mouffette de *Walt Disney* qui est amoureuse d'un chat noir orné d'une bande de peinture blanche sur le dos[23].

Bref, ici, des caractéristiques ne sont pas réunies sous un concept par une opération de synthèse ou un calcul statistique, mais plutôt en fonction des liens nomologiques, logiques, causaux, probalistiques, etc. En d'autres termes, un concept est une théorie à propos d'une classe d'entités; une structure représentationnelle qui encode des généralisations causales, fonctionnelles, nomologiques, à propos d'une classe[24].

Le fait que chaque concept regroupe une grande quantité d'informations provenant de plusieurs autres concepts rend impossible le cloisonnement de ces concepts dans des définitions rigides. De plus, les concepts conçus comme des théories apportent une explication concernant leur développement dans le temps: les changements des concepts/théories sont redevables aux mêmes mécanismes cognitifs qui sont responsables du changement des théories scientifiques.

2-Problème de la compétition entre les théories

L'objectif implicite de ces trois théories sur la nature des concepts est ultimement d'expliquer certaines compétences de la cognition supérieure[25]. En effet, en psychologie, comme évoqué précédemment, l'on soutient que les gens catégorisent et raisonnent de la façon qu'ils le font puisque ces processus utilisent des concepts qui possèdent toutes les mêmes propriétés. Il est donc essentiel d'expliquer ces propriétés pour atteindre l'objectif. Toutefois, la variété des théories sur la nature des concepts a soulevé une variété de types de comportements ou de phénomènes observés. Les trois théories proposent différentes explications de jugements

---

[23] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence.1999.

[24] Machery, *Concept is not a natural kind,*

[25] Machery, *Doing without concept*, 2009.

catégoriels et de raisonnements inductifs ou déductifs qu'un sujet cognitif peut faire. Par exemple, parfois un individu catégorise en comparant une représentation X avec une somme statistique de propriétés typiques d'un groupe de représentations mémorisées. Parfois, un individu catégorise en comparant une représentation X avec des membres *importants* d'un groupe de représentations mémorisées. Parfois, il catégorise si une représentation X a les propriétés qu'il s'attend à voir chez les membres d'un groupe de représentations en vertu des causes qui font que les membres sont ce qu'ils sont.

Il y a donc un débat entre ces théories, et ce, précisément parce que chacune d'elles prétend être la seule à décrire ce qu'est un concept. En effet, toujours selon Machery, la stratégie des théoriciens est de tenter de découvrir des propriétés de la cognition supérieure qui sont facilement explicables par leur théorie, mais en même temps inexplicables par les théories adverses. Cette compétition fait sens seulement si les chercheurs qui soutiennent une théorie supposent qu'il existe une seule théorie susceptible d'expliquer la nature des concepts. Si au contraire, l'on croyait que plusieurs schémas d'explication étaient équivalents, les théories porteraient sur différents types de choses, différents types de représentations mentales et, de la sorte, une découverte engendrée par la théorie E ne menacerait pas la théorie P. Mais ce n'est pas le cas, tous s'entendent sur le fait qu'il existe un seul type de choses qui réfère au concept de concept: une structure représentationnelle qui est retrouvée par défaut dans la mémoire à long terme lorsque quelqu'un raisonne ou catégorise et qui possède les propriétés x, y, z, ...

Or, le problème selon Machery (et d'autres)[26] est qu'aucune des théories formulées en psychologie ne décrit ni n'explique à elle seule tous les phénomènes cognitifs observés en psychologie et associés à l'utilisation de concepts. C'est-à-dire que ni la théorie des *exemples*, ni celle des *prototypes*, ni celle des *théories* ne parvient à

[26] *Machery, Weiskopf, Piccinini (voir bibliographie)*

prédire toutes les variantes de jugements catégoriels, ni toutes les variantes de raisonnements inductifs ou déductifs.

2.1 - Solution de l'hétérogénéité

Selon Machery[27], la classe des représentations mentales que l'on nomme «concept» dans les sciences de la cognition se divise plutôt en différentes classes de représentations mentales, en différentes sortes de *concepts* (*fundamental kinds*). Ainsi, il est vain de supposer qu'il existe des propriétés générales communes à tout ce que l'on nomme «concept» et qu'une théorie pourrait en faire la description.

Selon cette hypothèse, une même chose dans le monde est représentée distinctement par plusieurs types de concepts différents. Chacun de ces types de concepts qui réfèrent à la même chose peut être utilisé pour catégoriser différemment et raisonner différemment selon qu'il forme une structure statistique ou causale par exemple. Plus concrètement, un douanier a certainement un groupe d'*exemples* mémorisés de terroristes, peut-être Ben Laden, et il se sert sûrement des caractéristiques spécifiques des membres importants de ce groupe pour certains types de catégorisation, d'induction ou de déduction. Il a certainement aussi un *prototype* qui synthétise les caractéristiques typiques de ce qu'est un terroriste et il cherchera surement à le comparer avec certains voyageurs. Finalement, un douanier a certainement un ensemble de connaissances structurées de manière causale et nomologique concernant les terroristes, par exemple, il sait surement qu'ils peuvent dissimuler une arme.

Bref, si l'hypothèse de l'hétérogénéité s'avère véridique, la classe des concepts ne constitue pas une espèce naturelle. Il n'existe pas un type de représentations mentales qui sont retrouvées par défaut dans la mémoire à long terme lorsque

[27] *Machery, Doing without concept, 2009*

quelqu'un raisonne ou catégorise et qui possède les propriétés x, y, z, … Il existe plutôt des *exemples*, des *prototypes* et des *théories* qui permettent de catégoriser et de raisonner différemment sur une même chose et qui ont en fait peu en commun.

En effet, selon Machery, si l'on fait un survol de la recherche déployée depuis les années 70 concernant la catégorisation et l'induction, l'on découvre que certains phénomènes sont bien expliqués si les concepts mis à jour par quelques tâches expérimentales sont des prototypes, d'autres phénomènes sont bien expliqués si les concepts sont des exemples et d'autres phénomènes encore sont bien expliqués si les concepts sont compris comme des théories. Si l'on considère que les conditions expérimentales constituent la principale raison pour soutenir un type de concept plutôt qu'un autre type, ce dernier survol de la recherche fournit des arguments pour l'hypothèse de l'hétérogénéité.

Bref, puisque la psychologie a découvert empiriquement qu'il y avait différents processus cognitifs qui utilisaient des structures représentationnelles qui ont peu à voir entre elles, il est illusoire de vouloir les décrire en bloc comme des concepts. Ce sont différentes espèces. Et par conséquent, il faudrait considérer l'option d'éliminer le concept de concept qui ne réfère à rien d'empiriquement observable[28].

"the notion of concept is ill-suited to formulate scientifically relevant generalizations about the mind" (Machery, 2005, p. 464-5)

2.2 - Critique de l'éliminativisme à la Machery

Est-ce bien le cas que les différents types de représentations *conceptuelles* n'ont pas assez de propriétés communes pour que l'on puisse discuter de membres d'une

[28] Cette idée nécessite d'être davantage expliquée, ce qui se fera dans un travail subséquent.

même classe? Le concept de concept est-il vraiment inutile pour discuter des représentations utilisées dans la cognition supérieure? Nous n'en sommes pas certains. Plusieurs critiques de l'éliminativisme à la Machery sont possibles. D'abord, nous verrons avec Weiskopf dans la prochaine section que les différentes représentations qui remplissent la fonction des concepts ne peuvent pas être expliquées totalement de manière indépendante et isolée. Ces représentations forment plutôt un système conceptuel à propos duquel l'on peut appliquer plusieurs niveaux d'explication: un niveau général englobant les sortes de représentations et un niveau inférieur relevant les spécifications de chaque sorte de représentations.

2.2.1-Critique de Weiskopf[29]

Weiskopf soutient tout comme Machery qu'il existe une variété de représentations qui semblent satisfaire la fonction que l'on nomme *conceptualisation* en psychologie. En fait, selon lui, il y a au moins des *prototypes*, des *exemples*, des *modèles causaux,* comme pour Machery, et il y ajoute des *idéaux*[30].

Face à cette pluralité, Weiskopf n'opte pas pour éliminer la classe «concept» pour la diviser en différentes *espèces*. Il est vrai que ces *espèces* fournissent des explications plus fines et que l'on peut ainsi prédire des comportements précis relatifs à des tâches précises. Toutefois, il y a des propriétés que l'on ne peut expliquer qu'à un niveau supérieur. Par exemple, les représentations qui remplissent la fonction de concept sont toutes mentales, elles peuvent toutes être combinées dans

---

[29] Weiskopf, *The plurality of concepts,* 2008

[30] Les *idéaux* seront décrits dans le mémoire, mais très brièvement, les *idéaux* contiennent de l'information à propos des caractéristiques détenues par le meilleur exemple d'une catégorie et/ou sur les caractéristiques qu'il devrait détenir. Souvent, ils sont acquis en fonction des buts visés par un individu, plutôt que par une extraction à partir des perceptions. Dans le même sens, ils peuvent être acquis aussi par transmission culturelle.

des structures cognitives plus larges et elles ont toutes un rôle à jouer dans la catégorisation.

Machery rétorque que ces propriétés ne sont pas intéressantes[31] scientifiquement et surtout qu'elles ne sont pas le fruit de découvertes empiriques. Elles seraient plutôt des postulats analytiques.

Soit, Weiskopf soulève quatre autres propriétés générales (p.32) qui ne peuvent être expliquées au niveau des sous *espèces* et qui militent en faveur des concepts compris comme une classe.

1) Primo, lorsqu'elles effectuent une inférence du type *Lemieux est un boxeur québécois*, à *Lemieux est un boxeur et Lemieux est québécois,* les différentes sortes de concepts sont subsumées, à un niveau supérieur, par la forme logique ou syntaxique. D'autres auteurs tels que Sperber[32] et Piccinini[33] abondent dans le même sens.

2) Deuzio, il y a des propriétés qui émergent de la combinaison de différentes sortes de concepts et qui ne font pas partie des *sous-espèces* qui participent à cette combinaison. Weiskopf donne l'exemple du concept de *cuillère de bois* qui a la propriété d'être *large,* alors que ni les *cuillères*, ni le *bois* ne sont typiquement *large.*

3) Tertio, les différentes sortes de concepts partagent toutes les mêmes modes d'acquisition. Les prototypes sont souvent considérés comme une somme statistique extraite à partir d'*exemples* tirés de l'expérience. Les *théories* ou

---

[31] Machery, *Doing without concept*, 2009

[32] Sperber, *A pragmatic perspective on the evolution of language*, 2009

[33] Piccinini et Scott, *Splitting Concepts*, 2009

modèles causaux seraient aussi construits à partir d'exemples et à partir de prototypes. Il y aurait donc une interdépendance entre les sortes de concepts pour leur acquisition. De plus, les concepts partagent un autre mode, ils peuvent être transmis culturellement sans s'appuyer sur l'expérience. Par exemple, selon Weiskopf qui s'inspire de Putnam(p.35), en indiquant le sens d'un terme à quelqu'un, on lui transmet souvent le prototype associé à ce terme. Un *idéal* ou un *modèle causal* peuvent très bien aussi être transmis culturellement sans base perceptuelle.

4) Finalement, le processus qui gouverne la mémorisation et la récupération de l'information dans la mémoire à long terme doit pouvoir manipuler chaque sorte de concept, autrement il n'y a pas moyen d'expliquer le fait que le système conceptuel fonctionne de manière contextuelle. En effet, pour qu'il y ait une cohérence et pas de redondance entre les concepts durant l'encodage et la récupération, la mémoire à long terme doit permettre au sujet pensant de faire un lien entre ces concepts. Ce lien est tout simplement un lien d'identité dû au fait qu'ils réfèrent à la même chose. De plus ce lien, selon Weiskopf, est construit lors de l'acquisition qui suit une chaîne référentielle, confortant l'idée d'une interdépendance lors de l'acquisition (ci-haut). Ce mécanisme débuterait avec l'encodage de représentations exemplaires qui permettraient ensuite à un prototype d'être généré. Du fait que le prototype est généré grâce aux exemples comme input, il devrait être lié de manière coréférentielle à ces derniers. Lorsque que plusieurs prototypes sont générés, ils peuvent servir de passerelle pour acquérir d'autres concepts. Dans ce cas, l'information concernant le lien entre deux prototypes est encodée comme modèle causal dans la mémoire à long terme. De la sorte, le modèle causal, le prototype et l'exemple partageraient un lien d'identité; lien qui dépasse chacune de ses parties et qui ainsi assurerait une cohérence et une non-redondance.

Que retenir de ces critiques? Certes, Machery a raison de soutenir qu'il existe une multitude de types de représentations qui, isolément, ne satisfont pas l'ensemble des propriétés de la fonction de conceptualisation en psychologie. Toutefois, il est peut-être possible de regarder les mêmes structures représentationnelles à l'aide de niveaux d'explication différents, un plus précis pour observer les différentes représentations servant à regrouper les membres d'une catégorie et un autre, plus général, pour décrire les propriétés communes, conceptuelles. Suivant cette idée de Weiskopf, la conceptualisation est peut-être un travail (d'acquisition, de reconnaissance, de transmission) qui permet et utilise un complexe de représentations. C'est-à-dire qu'il y a probablement une interdépendance entre les sortes de représentations due à leur acquisition (3), que cette acquisition suit une chaîne référentielle (4) et ainsi, qu'il y a probablement un système conceptuel dont les propriétés sont différentes de celles de ses parties (les différentes représentations généralisantes). Maintenant, est-ce que ces propriétés sont suffisantes pour soutenir qu'il existe, à un certain niveau d'explication, des complexes représentationnels dans la mémoire qui sont composés de différentes représentations généralisantes? Si c'est le cas, Machery a peut-être bien saisi chaque ingrédient de la recette du concept de concept, mais il ne devrait pas s'étonner que ceux-ci, individuellement, n'en forment pas.

2.2.2-Concept implicite vs explicite

Les points (1) et (2) de la critique de Weiskopf touchent un autre point important : les concepts participent aux processus linguistiques. Il est de mise en psychologie que pour chaque processus cognitif, il y a un type de représentation ou du moins un type de structure mentale qui y participe. Par exemple, la perception de bas niveau et le raisonnement par l'image utiliseraient des images mentales. Les concepts eux participeraient aux processus de catégorisation et de raisonnement selon Machery.

Mais, selon Weiskopf, Piccinini et Scott, ils participent aussi aux processus linguistiques.

Piccinini et Scott[34] reprochent à Machery de ne pas discuter de ces processus (2006). Selon eux, on peut séparer les concepts en deux types, ceux qui nécessitent la compréhension et l'utilisation d'une langue naturelle et les autres. Les premiers sont observables seulement chez les humains capables de parler et les autres sont partagés par la plupart des animaux. Ces derniers sont utilisés dans des processus de discrimination, d'inférence non-linguistique et de catégorisation. Les *exemples*, *prototypes* et *théories* se résumeraient tous à ces concepts non-linguistiques. Les premiers sont utilisés dans des processus de compréhension de langage, d'inférence linguistique et de combinaison lexicale. Piccinini et Scott nomment ces deux types de concepts : implicite et explicite.

Nous n'entrerons pas dans les détails de l'argumentation de Piccinini ici, retenons seulement 3 types d'arguments qui soutiennent l'idée de concept explicite. Primo, il y a des arguments neurobiologiques qui démontrent des zones cérébrales reliées à la catégorisation essentiellement linguistique. Deuzio, il y a des arguments concernant le lien entre les concepts et les processus syntaxiques. Ici les structures sémantiques qui sont utilisées dans les processus linguistiques doivent avoir le format requis pour être manipulées syntaxiquement. En l'occurrence les concepts doivent jouer le rôle du nom, verbe, adjectif, etc. Ils doivent incarner de l'information syntaxique dans leur structure même. Tertio, certaines choses ne peuvent être représentées que par des concepts linguistiques : les choses inobservables comme les trous noirs ou les atomes, les choses non-réelles comme les super héros, les choses abstraites comme la justice ou la vérité, etc.

---

[34] Piccinini et Scott, *Splitting Concepts*, 2009

Bref, Piccinini montre que Machery s'est facilité la tâche en réduisant au minimum la nature des concepts à fin de les éliminer. C'est aussi ce que semble soutenir d'autres auteurs comme Manrique[35] ou P. Poirier et Beaulac. Ces derniers écrivent que Machery développe une *attitude négative, à contre-courant de la tendance actuelle en sciences cognitives, à l'égard des théories à processus duel*[36] (de type implicite/explicite, donc plus complexe).

Poirier et Beaulac nous rappellent qu'une autre façon de distinguer différents niveaux de complexité entre les types de représentations généralisantes est de distinguer les concepts utilisés ou réalisés par des processus automatiques et ceux utilisés ou réalisés par des processus réfléchis. Par exemple, le premier type de représentations (type 1), qui regrouperait les *espèces* de Machery, serait utilisé dans des processus automatiques, alors que le deuxième type de représentations (type 2), utilisé dans des processus réfléchis, regrouperait des concepts plus classiques nécessitant du langage, des définitions :

...des corps d'information servant à regrouper les membres d'une catégorie peuvent également être construits et acquis par des processus réfléchis. Typiquement, ces corps d'information prennent la forme de définitions transmises verbalement et appliquées par la remémoration de l'énoncé verbal de la définition[37].

---

[35] Fernando Martínez Manrique, *Dual minds, dual concepts?* Universidad de Granada (2010)

[36] Pierre Poirier et Beaulac et Guillaume Beaulac, *Le véritable retour des définitions*, UQAM (2010)

[37] Pierre Poirier et Beaulac et Guillaume Beaulac, ***Le véritable retour des définitions***, Université du Québec à Montréal (2010)

2.3- Notre position dans le débat

Nous sommes d'accord avec l'éliminativisme de Machery concernant les *exemples*, les *prototypes* et les *théories*. En effet, aucune d'elles ne décrit ni n'explique à elle seule tous les phénomènes cognitifs observés en psychologie (associés à l'utilisation de concepts). Certains, comme Piccinini, en concluent qu'il n'y a pas qu'un seul type de concept, mais une hétérogénéité de représentations qui partagent certaines propriétés et qui par conséquent sont subsumées sous la classe «concept». Mais, il n'existe pas différentes sous-espèces de concepts. Selon Machery parce que leurs contenus et les processus qui les utilisent sont trop différents l'une de l'autre pour être regroupées sous une classe, ce que nous refusons; selon nous, il n'existe pas différentes sous-espèces de concepts précisément parce qu'aucune de ces représentations en psychologie n'explique le phénomène de conceptualisation si l'on tient compte tant de l'architecture qui la soutient que des contenus qu'elle traite et des processus qui lui sont associés.

En effet, les trois espèces de représentations en psychologie suivent une chaine référentielle lors de leur acquisition, ils correspondent en fait à trois étapes d'un même apprentissage hiérarchisé (Weiskopf). Par conséquent, nous soutenons l'hypothèse qu'il existe un seul type de structure dans la mémoire à long terme (du cortex)[38] : un complexe de représentations superposées, duquel différents types d'informations peuvent être utilisés, selon le contexte, pour des traitements futurs. C'est-à-dire, qu'il serait surprenant qu'un individu stocke un *exemple* sur une population de neurones, un *prototype* référent à la même chose sur une autre et une *théorie* coréférentielle sur une troisième. Cette possibilité soulève d'une part le problème de la cohérence et de la redondance lors de l'utilisation contextualisée des concepts (Weiskopf); et d'autre part, l'utilisation de différentes populations de neurones pour référer à une même chose ne semble pas rentable biologiquement. Il

[38] James L. McClelland, *Why there are Complementary Learning Systeme in the Hippocampus and Neocortex*, 1994

nous semble qu'il n'existe pas différents concepts, mais différents niveaux de représentations, de complexités différentes, qui forment un complexe conceptuel.

C'est pour cette raison que nous sommes d'accord avec l'élimination du terme « concept » pour référer à chacune des (niveaux de) représentations implicites/type1, mais de le garder lorsqu'il réfère à une structure regroupant les trois à la fois.

De plus, nous sommes d'accord aussi avec Piccinini et Scott lorsqu'ils affirment que Machery, en ne discutant pas des propriétés propres aux concepts explicites, ne démontre pas pourquoi l'on devrait les éliminer. Par conséquent nous soutenons l'idée qu'une théorie des concepts devrait expliquer leur utilisation dans des processus de niveaux supérieurs (explicite/type2) ce que les représentations de Machery, isolément, n'expliquent pas. Nous suggérons donc que les structures représentationnelles de type2 soient peut-être en fait des complexes de représentations (de type1) susceptibles d'être utilisés dans les processus linguistiques, réfléchis et/ou de former des définitions. C'est ce que nous verrons maintenant.

Tableau 1.2 : Différentes positions dans le débat sur l'hétérogénéité des concepts

Machery	3 types de représentations	---
Piccinini	Concept implicite	Concept explicite
Théorie duelle (Poirier et Beaulac)	Concept type 1	Concept type 2
Weiskopf	3 sous-classes de concepts	Concept (classe)
Brunetta	3 types de représentations non conceptuelles	Complexe conceptuel (explicite et réfléchi)

3-Concept : entre philosophie et sciences cognitives

Nous cherchons à fournir une explication de système conceptuel dans toute sa richesse. Nous cherchons d'une part une définition qui intègre toutes les fonctions *implicites* décrites par les *exemples*, les *prototypes* et les *théories.* En effet, nous croyons que ces représentations généralisantes sont les ingrédients de base pour réaliser de véritables concepts explicites et/ou de type 2. D'autre part, nous cherchons une définition qui intègre les fonctions *explicites* reliées aux compétences linguistiques (compréhension de langage, inférence linguistique et combinaison lexicale) et/ou aux processus réfléchis en général.

La critique de Piccinini et celle de Poirier et Beaulac nous invitent à penser que la solution est peut-être à chercher en posant un regard plus large, voir philosophique, sur les théories psychologiques, par exemple en rétablissant certaines propriétés des concepts véhiculées traditionnellement en philosophie comme la structure définitionnelle. Dans ce sens, pour nous sortir des œillères de la psychologie, nous cherchons une définition plus générale qui puisse regrouper certains critères traditionnels en philosophie et les fonctions proposées par les théories ci-dessus.

Critères pour les représentations implicites

Si l'on récupère une théorie plus classique et que l'on veut du même coup satisfaire les propriétés importantes soulevées par la psychologie, on peut , primo, soulever le fait que les concepts sont une synthèse de représentations singulières sous une représentation générale; qu'ils sont une opération de classification des choses communes; ensuite qu'ils sont une opération de généralisation dans la mesure où elles concernent une multitude d'occurrences; et, que la conceptualisation est un processus de stabilisation du flux perceptuel[39]. C'est ce que nous avons nommé les

[39] Meunier J-G, *Trois types de représentations*. Publication du LANCI. UQAM. 2002. www.unites.uqam.ca/lanci

critères endogènes en philosophie et qui peuvent répondre aux fonctions des représentations implicites en psychologie.

Critère pour les concepts explicites

Deuzio, concernant les critères exogènes en philosophie, la conceptualisation est une opération de représentation qui n'est pas appliquée aux objets externes eux-mêmes, ou répondant à des stimuli directement, mais à d'autres représentations. Précisément, ce sont des représentations sur des représentations déjà classifiantes. De plus, et surtout, les concepts participent à la cognition supérieure. Les concepts explicites en psychologie répondent déjà à ce critère. En effet, comme le souligne Piccinini, on doit concevoir des concepts qui participent aux opérations linguistiques (cognition supérieure). Et, suivant Poirier et Beaulac, les concepts utilisés dans des processus réfléchis (cognition supérieure) peuvent prendre la forme de définitions.

Tableau 1.3 : Critères pour notre définition de concept : entre philosophie et psychologie

Critères exogènes et/ou implicites	Critères exogènes et/ou explicites
Opération de classification	Représentation sur une représentation
Ces classes sont généralisantes et intégrantes	Participent aux opérations supérieures
Stabilisation du flux perceptuel	Peuvent prendre la forme de définitions

Maintenant, une autre façon de prendre un recul face à la psychologie est de regarder les niveaux d'analyse plus larges en sciences cognitives. En effet, si en psychologie on part souvent d'observations sur l'utilisation de contenu conceptuel pour en inférer les sortes de véhicules conceptuels, une explication complète en sciences cognitives doit intégrer aussi un modèle cognitif (physique) dont le fonctionnement explique

les observations comportementales[40]. Or, existe-t-il un modèle de système cognitif qui peut réaliser le travail décrit par cette définition et expliquer comment des structures représentationnelles explicites/type2 pourraient-être construites comme des complexes de représentations implicites/type1? Nous croyons que les modèles connexionnistes classiques peuvent expliquer de manière concrète et mécanique, du moins par des moyens informatiques, plusieurs des caractéristiques d'un tel système.

---

[40] Adrian Cussins, *The Connectionist Construction of Concepts*, in *The Philosophy of Artificial Intelligence*, edited by Margaret Boden, in the *Oxford Readings in Philosophy* Series (Oxford University Press, 1990), pp. 368 - 440.

CHAPITRE II

RÉSEAU DE NEURONES

1- Réseau de neurones comme modèle de système conceptuel internaliste

Modèle de réseau de neurones de Churchland. Notre but ici est de chercher à comprendre comment les modèles connexionnistes de la cognition répondent au problème de la conceptualisation. Nous définissons cette opération dans le premier chapitre en partie comme un processus de généralisation de l'information perçue par un individu. Pour un modèle connexionniste de la catégorisation, ce processus débute par de multiples inputs toujours changeants sur un capteur et tend vers une stabilisation, vers des traits généraux et durables, c'est-à-dire vers un possible concept. Nous décrirons comment un réseau représente les traits généraux et durables *du* monde par une configuration stable de ses poids de connexions synaptiques, alors qu'il représente les inputs singuliers et éphémères par une configuration transitoire de niveaux d'activation.

Tableau 2.1

Représentation des inputs singuliers et éphémères	Représentation des traits généraux et durables
Vecteur de niveaux d'activation	Vecteur de connexion synaptique

Transitoire, change rapidement	Très stable, change lentement

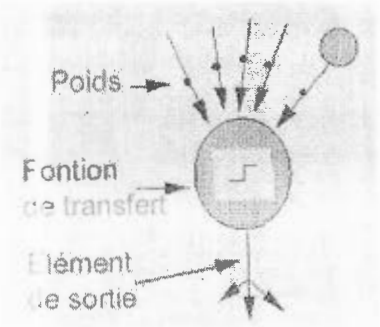
Ce processus de conceptualisation sera alors un travail de transmission et d'intégration des vecteurs de niveaux d'activation par les configurations de poids de connexions synaptiques des différentes couches de cellules que possède un agent. C'est-à-dire que les configurations de poids de connexions synaptiques des différentes couches sélectionnent, comme un crible sélectionne, une partie des vecteurs de niveaux d'activation situés dans les couches en amont et suppriment les autres parties[41]. Le vecteur du niveau d'activation du capteur n'est pas le même qui aboutit aux dernières couches de neurones. C'est cette transformation qui nous intéresse. Comment extraire les propriétés essentielles d'un niveau d'activation d'un capteur et non celles qui sont accidentelles?

1.1 Encodage : connexionnisme

Les modèles mathématiques de réseaux de neurones artificiels représentent des réseaux de processeurs élémentaires interconnectés fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Généralement, on calcule d'abord l'état d'activation du neurone par la somme des poids en entrées ; ensuite on compare cette somme au seuil de la fonction de transfert du neurone, qui déterminera la valeur de la sortie[42]. Nous verrons en détail ce traitement neuronal plus loin, mais pour l'instant retenons que chaque neurone d'un réseau reçoit un input qui détermine son niveau d'activation et peut émettre un output selon la fonction de transfert.

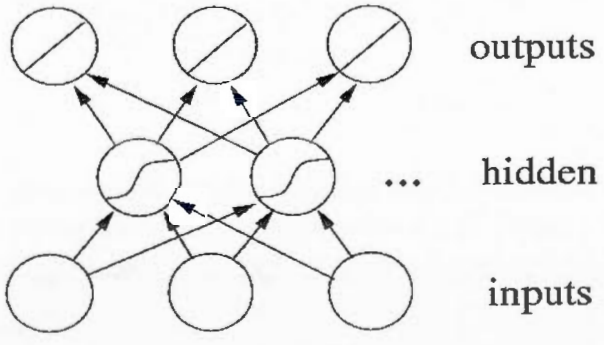
Figure 2..1: modèle de neurone artificiel

[41] En fait, nous le verrons, ça fait plus que ça : le réseau déforme l'espace d'activation.
[42] Toutes les fonctions de transfert n'ont pas un seuil. Une fonction logistique, par exemple, n'a pas vraiment de seuil, même si elle est a un point d'inflexion qui ressemble à un seuil.



À l'intérieur de certains réseaux, plusieurs couches hiérarchiques de neurones peuvent former des sous-réseaux en soi[43] : ces modèles sont dits multicouches, dont le plus classique est le modèle *feedfoward*. L'architecture générale d'un tel réseau reproduit un schéma d'entrée/sortie proche de celui d'un neurone individuel, c'est-à-dire qu'il possède une couche d'entrée, une de sortie et souvent des couches cachées entre les deux.

Figure 2.2: modèle de réseau de neurones multi-couche



Dans un modèle connexionniste multicouche, la première impression d'un stimulus extérieur sur le système est subie par la couche d'entrée. Les cellules de la couche

[43] Claude Touzet, *Introduction au modèle connexionniste*. 1992. dans google.

d'entrée reçoivent leurs intrants non pas d'autres cellules du réseau, mais de l'environnement externe au réseau.

1.2 Vecteur d'activation

Chez l'homme par exemple, une des couches d'entrée est la langue. Selon Churchland[44], la langue une fois modélisée forme une matrice de cellules sensorielles divisées en quatre types: les cellules sensibles au sucré, à l'acide, au salé et à l'amer. Chacune de ces cellules réceptrices réagit ou s'active à des niveaux différents selon les intrants. Churchland fait l'hypothèse qu'il y a environ 10 niveaux distincts par type de cellule. Si on présente un jalapeño au capteur gustatif par exemple, chaque type de cellules réceptrices s'active à un certain niveau. En l'occurrence, le sucré au niveau 3 sur l'échelle de 10, le salé à 4, l'acide à 9 et l'amer à 4. Pour le jalapeño, la représentation numérique de l'activité de la langue (selon les 4 types) peut être le vecteur : v3494.

La signature du jalapeño sur la matrice gustative ne dépend pas d'un type de cellule réceptrice en particulier, c'est la configuration collective des quatre types de cellules qui importe. Cette signature du stimulus est un pattern ou un vecteur d'activation distribué sur toute la couche d'entrée, de manière globale. Un piment n'est pas représenté uniquement par les cellules acides et encore moins par une seule cellule de manière locale.

Ainsi, lorsqu'un piment fort ne produit pas le même pattern d'activation sur la langue qu'une bouchée de crème glacée par exemple, la différence s'exprime entre les patterns ou vecteurs de manière distribuée: v3494 et v8442[45].

---

[44] Churchland Paul, *The engine of reason, the seat of the soul: a philosophical journey into brain.* First MIT Press paperback edition, 1996

[45] Notons toutefois que l'exemple de Churchland est un peu simpliste. En réalité le goût piquant du jalapeno comme le goût de la crème glacée ne résultent pas uniquement de senseurs gustatifs. Le

1.3 Combinatoire

Un tel système distribué est très sensible. Si chaque type de cellule peut s'activer à 10 différents niveaux, comme le dit Churcland, alors les quatre types ensemble (10*10) peuvent être affectés de 10 000 manières différentes[46]. Cette stratégie d'encodage qui utilise des configurations de collection de cellules est donc une stratégie combinatoire.

1.4 Encodage superposé

Un autre avantage pour le modèle de l'encodage par combinatoire est la superposition de l'encodage. Puisque toutes les cellules d'une matrice sensorielle sont activées globalement en présence de stimuli externes, différents stimuli singuliers sont encodés par le même ensemble de cellules. Par conséquent, deux signatures (vecteurs d'activation) sur une même matrice provenant de deux stimuli quantitativement différents sont absolument superposées si et seulement si les patterns ou les vecteurs de l'encodage sont identiques. L'avantage de cette superposition se trouve dans la proximité des signatures, des vecteurs; on dira que deux vecteurs sont plus ou moins superposés sur une matrice sensorielle. Déjà, au niveau sensoriel, il y a une possibilité d'un éventuel travail de comparaison des similarités en aval du système ou pour un agent extérieur.

---

piquant dépend d'un senseur pour la capsicine, la crème glacée pour le gras de la crème. Maintenant, notre intention est de décrire un modèle connexionniste général et non spécifique au goût. Bref, cette approximation est suffisante pour notre propos

[46] Maintenant, Churchland se sert de la sensibilité du modèle de la langue pour soutenir que les mots du langage ne sont pas aussi précis que la sensation peut l'être et par conséquent que le langage est peut-être moins important qu'on pensait. En fait, si l'information sur un capteur précède toute perception, alors son argument ne tient plus puisqu'on n'a pas conscience de cette information directement.

2-Traitement de l'encodage ou perception

Selon le modèle de Churchland, ces représentations vectorielles qui émergent des capteurs sont codées et détectées par d'autres types de cellules qui se trouvent non pas en périphérie sensorielle du système cognitif, mais plus loin dans la chaîne des couches de cellules.

Mais avant toute chose, précisons que lorsqu'elles sont reconnues par les couches en aval, les représentations vectorielles singulières qui émergent des couches d'entrée doivent être synthétisées et non simplement ré-encodées telles quelles. On n'encode pas davantage les milliards de stimuli qui affectent nos sens. C'est-à-dire, que s'il y a un processus d'encodage d'information en aval sur des informations déjà encodées en amont, alors ce deuxième encodage ne doit pas être qu'une simple reproduction de l'état de la matrice du capteur que l'on stockerait à long terme. L'esprit n'est pas une accumulation de représentations particulières au fil de l'expérience. Pour nous en convaincre, pensons qu'il est difficile de se souvenir du vecteur d'activation provenant de notre langue le 3 mars 2007 à 15h45m12s ou encore plus le 3 mars 1979 à la même heure, alors qu'il est facile d'affirmer sans trop de conséquence qu'on a mangé de la nourriture ces deux jours-là. La raison est que c'est ce que l'on fait en *général* à tous les jours. Ici, il n'y a pas seulement encodage et stockage d'informations singulières, il y a utilisation d'outils formels synthétiques et généraux.

Maintenant, comment le connexionnisme explique-t-il qu'un agent cognitif puisse synthétiser des vecteurs d'activation et même circonscrire une classe? Quelles sont les fonctions qui gèrent le travail de synthétisation et de classification?

2.1 Ébauche du prototype

D'abord, nous disions plus haut que la distributivité et la superposition de l'encodage permettent à un ensemble d'objets de partager les ressources disponibles sur un capteur et nous permettent d'envisager un travail de comparaison des similarités. C'est-à-dire qu'il y a similarité entre différentes signatures de sensibles sur le capteur qui permettent une éventuelle comparaison.

Maintenant, cette tendance des réseaux à représenter des similarités par l'utilisation de ressources communes permet à des couches en aval ou à un agent cognitif supérieur de *calculer*[47] une *moyenne des représentations* sur le capteur. En d'autres mots, l'utilisation de ressources communes permet l'extraction d'une représentation prototypique en aval. L'exemple de Churchland est celui d'une représentation d'un visage prototypique qui serait la moyenne d'un échantillon de visages divers encodés par la couche d'entrée. Ici, puisque les représentations sont des vecteurs numériques, il est très facile de calculer une moyenne, un prototype. Par exemple, pensons à une matrice d'entrée possédant 20 types de cellules qui seraient chacune sensible à une propriété particulière. Ces 20 types correspondraient sur le vecteur à 20 dimensions (V=d1, d2...d20) à qui l'on présenterait 100 visages. Imaginons que chacune des 20 dimensions code un aspect du visage en s'activant à 5 différents niveaux. Ainsi, pour calculer la moyenne de 100 vecteurs d'activation singuliers (visages), il suffit d'additionner pour chacune des 20 dimensions les 100 niveaux d'activation (entre 1 et 5) et de les diviser par 100. Au bout du compte, on obtient les 20 dimensions moyennes du vecteur moyen qui codent le visage prototypique.

À quoi peut bien servir un prototype? Justement à encoder de manière synthétisée, classée et généralisante. Selon plusieurs modèles connexionnistes, un prototype est une fabrication du système qui sert d'exemple typique pour comparer et classer les choses. Ainsi, on dit du moineau qu'il est plus typique pour la classe des oiseaux

[47] En fait, les couches en aval se comportent comme s'il calculait une moyenne mais ne le font pas (pour prendre une vieille distinction, elles se comportent comme si elles suivaient une règle mais n'en suivent pas une.

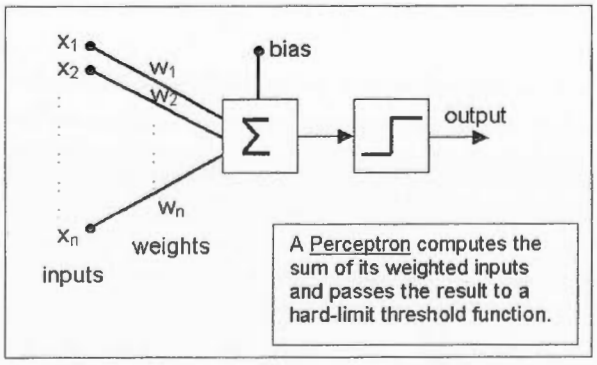
que ne l'est le pingouin, parce que le vecteur du moineau se rapproche numériquement plus du prototype d'oiseau que le fait celui du pingouin. Plus besoin pour un agent d'encoder tout ce qui se présente à lui, l'agent fabrique ou configure un prototype et s'en sert pour intégrer et structurer le flux de perceptions, il calcule la différence entre les vecteurs singuliers et le vecteur prototypique. Un système connexionniste peut même définir des zones prototypiques ou des intervalles vectoriels qui intégreraient les vecteurs singuliers.

Soit, mais comment une couche cachée calcule ou fabrique, dans les faits, de telles représentations?

2.2 Perceptron

Certaines cellules de couche cachée apprennent à détecter ou à s'activer en présence de certains types de vecteurs d'activation à l'entrée. En fait, la plupart des cellules de réseau connexionniste s'activent lorsqu'elles reçoivent en entrée un niveau d'activation supérieur à leur seuil de transfert. Le Perceptron est l'exemple typique dans ce cas-ci :

Figure 2-3: computation d'un perceptron



Le Perceptron reçoit des potentiels d'actions (pour simplifier disons qu'ils sont binaires : excitateurs 1 ou inhibiteurs -1) de cellules X en amont, chaque connexion entre un X et le Perceptron est pondérée par un poids W. Ensuite, le neurone calcule la somme $\sum$ des grandeurs en entrée, pondérée par les poids de connexions, et la compare à une fonction de transfert régulée par un seuil S limite. La sortie O est binaire, c'est-à-dire que le Perceptron divise les entrées en deux parties.

2.3 «Patron préféré»

Dans le vocabulaire de Churchland[48], le seuil d'activation d'une cellule est un «patron d'entrée préféré»[49] d'une cellule. Différentes cellules s'activent selon différents patrons de poids à l'entrée. Prenons l'exemple d'une cellule qui reçoit des poids à partir de 9 autres cellules en amont. Disons que ces 9 entrées proviennent plus précisément d'un capteur d'images tel qu'une matrice de caméra numérique et que cette matrice est disposée en 3 rangées de cellules binaires par 3 colonnes de mêmes cellules binaires (9 cellules au total pour 2^9 vecteurs possibles). Maintenant, quand il y a un vecteur $X=x_1...x_n$ tel que représenté dans le tableau suivant qui prend la forme imagée d'un T, la cellule en aval s'active. Car cette cellule en particulier possède une configuration de poids de connexions $W=w_1...w_n$ qui *reconnaît* ce vecteur d'activation en entrée[50].

---

[49] Churchland Paul, *The engine of reason, the seat of the soul: a philosophical journey into brain*. First MIT Press paperback edition, 1996 (p 53)

[50] Ici nous ne sommes pas certain si le vecteur qui active la cellule réceptrice peut être une somme pondéré comparée à un seuil ou plutôt un patron de poids qui serait associé à une cellule en aval par une autre règle telle que celle de Hebb, c'est-à-dire une règle d' habituation?

Figure 2.4 : Patron d'activation 1

1	1	1
0	1	0
0	1	0

Bref, pour Churchland les cellules cibles *préfèrent* certains patrons d'entrée. Ce modèle fournit de bons outils pour notre problématique : ce type de cellule réceptrice synthétise et sélectionne de l'information inscrite en amont en adoptant (ou en recevant) une configuration de poids de connexions plus qu'une autre. En effet, dans notre exemple, l'information de 9 cellules est synthétisée par une seule cellule et seulement un groupe restreint de patrons d'activation est disposé à activer une seule cellule réceptrice en fonction de son patron de poids (et de son seuil ?).

Soit, mais est-ce suffisant pour que l'on puisse vraiment dire que ce type de cellules détectrices encode des prototypes tels que définis plus haut? Car, une cellule individuelle détecte différents sensibles en autant que ceux-ci produisent un vecteur d'activation en entrée plus ou moins proche, mais il n'y a pas encore de comparaison possible entre les similarités d'objets différents tels qu'un moineau et un pingouin. Une seule cellule ne peut pas classer positivement les deux classes d'animaux (sans que le système ignore l'une des deux).

2.4 Réseautage des stimuli préférés

Ici, la question est comment des cellules d'une couche cachée peuvent-elles grouper des vecteurs d'entrée différents qui partagent certaines similarités telles que les vecteurs hypothétiques associés aux moineaux et aux pingouins?

La réponse telle qu'Andy Clark[51] la propose est que dans les modèles connexionnistes multicouches, les couches cachées ne détectent pas les signatures émergeantes des capteurs de manière locale, mais de manière distribuée, combinée et superposée, comme le font les couches en entrée pour les sensibles. Dans ce contexte, un prototype n'est pas codé par une cellule en aval, mais par un réseau en entier. Chaque cellule du réseau en aval peut détecter des *features* particuliers qui seront combinés par le réseau pour former des prototypes ou autres représentations synthétiques. C'est en s'interconnectant que certaines cellules détectrices de *features* communs pourront classer les vecteurs d'activation en amont à un niveau de généralité plus élevé encore qu'avec les simples stimuli préférés.

Par conséquent, la moyenne vectorielle qui est extraite des vecteurs d'activation du capteur est opérée par l'ensemble des cellules de la couche cachée. Clark décrit cette extraction de la moyenne par le réseau en deux principales étapes. Primo, il décrit une tendance qu'il nomme «highly marked» proposant que les poids et seuils des connexions entre les cellules de la couche cachée qui encodent les *features* communs en entrée tendent à être très stables. Car, les connexions seraient modelées par ce qui se présente le plus souvent au réseau, en l'occurrence des propriétés qui sont communes à plusieurs intrants[52]. Deuzio, les choses présentées à un agent possèdent souvent plusieurs propriétés qui sont toujours associées à cette présentation. Ainsi, différentes propriétés apparaissent souvent ensemble. Clark nomme cette tendance «mutually associated».

Bref, c'est en réseau que le groupement des similarités s'effectue. Mais comment le réseau détecte ces nouveaux vecteurs encore plus synthétiques ?

---

[51] Clark A., *Associative engine*. MIT Press. 1993

[52] Il est question ici d'apprentissage du réseau, ce que nous verrons plus en détail dans la prochaine section.

3. Conceptualisation (ou non ?)

3.1 Couche de sortie

Jusqu'ici, les cellules des couches cachées sont sensibles à leurs stimuli préférés, elles détectent certaines propriétés communes (P, Q, R) à différents vecteurs d'activation en entrée, ces propriétés synthétiques sont ensuite utilisées dans la combinaison de différentes représentations plus complexes encore (P+Q+R= C), qui se rapprochent plus des sortes de choses que nous connaissons ou sommes conscients.

Maintenant, comme les cellules appartenant au capteur, les cellules de la couche cachée ne forment individuellement que des parties implicites de la configuration globale de leur couche, ainsi, les représentations complexes ou les combinaisons de stimuli préférés qui émergent de l'activité globale de la couche cachée ne sont pas détectées par ces mêmes cellules. Ces représentations complexes sont stockées et détectées par d'autres types de cellules qui se trouvent plus loin dans la chaîne des couches: pour un système à trois couches, ce sont des cellules de sortie[53] qui détectent ces représentations.

Les cellules de la couche de sortie détectent habituellement des sortes de choses pour lesquelles le système est conçu. Par exemple, un système qui est conçu pour détecter des mines antipersonnelles possèdent au moins une cellule en sortie qui s'active lorsque le traitement en amont fait émerger un vecteur d'activité *associé* aux mines. (L'activation de cette cellule peut causer ensuite des réactions motrices ou

[53] Notons que pour le système nerveux de l'homme, il existe plusieurs dizaines ou centaines de couches entre un organe sensoriel et la dernière couche du cortex, qui d'ailleurs seraient peut-être injustement nommées couche de sortie?

activer un symbole pour un contrôleur…). D'autres réseaux peuvent servir à encoder et détecter des fruits, des missiles, des chiens, des automobiles, des voleurs, etc. Pour un système artificiel, c'est le programmeur qui associe une classe à une sortie, pour les systèmes biologiques, qui possèderaient plusieurs modules de détection[54], c'est Dieu ou la sélection naturelle ou la culture ou des extraterrestres ou nos parents qui font l'association entre une sortie et une sorte de chose. Mais peu importe ici qui ou quoi décide d'une classe à associer à une cellule en sortie, ce qui importe pour l'instant est comment un système connexionniste *apprend* à associer une classe donnée et l'activité d'une cellule de sortie.

3.2 Apprentissage du réseau

Le but de l'apprentissage est de transformer chacun des vecteurs d'activités possibles en entrée en un vecteur de sortie discriminatoire approprié à cette entrée: produire un vecteur de sortie qui organise les vecteurs d'entrée en classes signifiantes pour le système lui-même ou pour le concepteur ou pour un observateur quelconque.

L'exemple de Churchland est un système qui organise des visages singuliers en entrée en quelques classes synthétiques. Ici, la couche d'entrée est une matrice carrée de 64 cellules par 64 (4096 unités) qui captent les exemples de visages présentés au réseau; ensuite la couche cachée possède 80 cellules qui synthétisent l'information en amont en répondant à leurs stimuli préférés; et la couche de sortie possède 9 cellules qui représentent les vecteurs implicites qui émergent de la couche cachée en quatre classes : A) une première cellule détecte si le vecteur représente un visage (1) ou non (0), B) une deuxième cellule détecte si le vecteur représente un homme (1) ou non (0), C) une troisième détecte la présence d'une femme (1) ou non

[54] Si l'on accepte la thèses multimodale populaire dans les sciences cognitives, voir Pinker

(0), et D) les cinq autres cellules encodent l'identité des visages selon un vecteur de sortie à cinq dimensions possédant chacune une dizaine de niveaux d'activation.

Avant de pouvoir discriminer les entrées en quatre classes, le système est complètement ignorant. Chacun des milliers de poids de connexions des cellules du système est aléatoire. Le réseau en entier forme un désordre insignifiant. La première étape est opérée par l'expérimentateur, en l'occurrence il détermine un type de sortie souhaitée en chargeant les cellules de sortie d'une signification. C'est-à-dire que l'expérimentateur associe une classe de choses à détecter à une cellule de sortie (généralement prévue avant même de construire le réseau comme tel).

Deuxième étape, l'expérimentateur définit un vecteur de sortie théorique souhaité en fonction d'une image présentée en entrée. Cette étape est spécifique de l'apprentissage supervisé[55], elle détermine la tâche à apprendre.

Troisième étape, il présente l'image X au système et il obtient un premier résultat erroné encodé, par exemple, par le vecteur V=(0.23 ; 0.80 ; 0.39 ; (0.20, 0.30, 0.19, 0.66, 0.96)).

Quatrième étape, l'expérimentateur soustrait ce premier vecteur erroné au vecteur souhaité. La différence obtenue est égale à un vecteur représentant *l'erreur du système*. S'il fait une moyenne des 9 dimensions, il obtient un nombre correspondant à la moyenne des erreurs : en occurrence 0.7158.

Pour corriger l'erreur, l'expérimentateur maintient tous les poids de connexions qui pondèrent le potentiel d'activité, sauf un qu'il modifie.

---

[55] Il existe une autre forme d'apprentissage qui est non-supervisé, celle-ci ne vise pas à enseigner une tâche à accomplir au réseau, mais mise plutôt sur des lois de corrélation ou d'habitus de type Hebbien. Voir Poirier P.

Ensuite, il regarde l'effet sur la moyenne des erreurs. S'il n'y a pas de changement, alors il laisse tout comme ça. S'il y a une hausse ou une baisse de la moyenne des erreurs, alors il ajuste le poids de la connexion pour que l'erreur se rapproche le plus du vecteur souhaité. Il répète la même procédure avec tous les poids du réseau. Celui-ci change petit à petit sa configuration de poids de connexions pour se rapprocher du vecteur souhaité par *l'apprenant*, comme aime le dire Churchland. À un certain moment, l'écart entre le vecteur souhaité et celui du système se stabilise.

Ce procédé d'apprentissage par ajustement des poids de connexions, appelé aussi rétropropagation, est répété pour chaque exemple faisant partie de l'ensemble d'apprentissage. Par exemple, le modèle de Churchland a été exposé à une quarantaine de visages lors de sa formation (la moitié était des hommes, l'autre des femmes).

3.3 Qu'est-ce qu'un réseau connaît réellement?

Après l'apprentissage, les poids de connexions du réseau sont configurés de telle sorte que la couche cachée n'est sensible qu'aux patrons de visages sur le capteur ou presque. Maintenant, les configurations d'activités des cellules de ces configurations de connexions représentent des visages implicites différents qui se ramènent tous à des vecteurs singuliers d'activation. Dans l'exemple de Churchland, ce sont des vecteurs possédant 80 dimensions puisque la couche cachée possède 80 cellules codant une propriété chacune (notons que chaque dimension synthétise déjà un vecteur de 4096 dimensions en entrée). Si l'on compare les vecteurs des 40 exemples d'apprentissage entre eux, on remarque des similarités dans l'encodage des visages féminins que l'on ne retrouve pas dans le cas de l'encodage des visages masculins, et vis versa. De plus, il est possible d'extraire un prototype des visages de femmes (et un des visages d'hommes) et de calculer la différence ou l'écart entre le prototype et un exemple singulier. Ce qui fait dire à Churchland qu'il y a des

espaces vectoriels ou des espaces d'activation des cellules dans la couche cachée qui se forment lors de l'apprentissage.

Le même processus de partition de l'espace d'activation des cellules est applicable aux sous-classes telles que le sexe et l'identité dans l'exemple utilisé ici. Et c'est évidemment ces partitions que les cellules de la couche de sortie détectent pour former plus explicitement des classes.

En plus d'intégrer et de classer les vecteurs d'activation, le réseau a acquis une certaine capacité de généralisation. C'est-à-dire que le réseau n'a pas simplement mémorisé les exemples d'apprentissage, il est apte à reconnaître de nouveaux visages (jamais perçus) comme des «visages» et ce avec une note de 100%. Il peut même déterminer le sexe dans 98% des cas (est-ce que les nouveaux exemples respectent des paramètres précis déterminés par le chercheur ? Est-ce que le système traite seulement des représentations dont le chercheur sait qu'il va reconnaitre?). Autre aspect intéressant, le réseau est capable de reconnaitre des visages familiers auxquels on a brouillé 1/5 de la surface. Bref, les classes formées lors de l'apprentissage peuvent intégrer le flux de perception qui se présente au capteur.

Il n'en faudra pas davantage pour que Churchland nous fournisse une réponse directe à la question de notre travail :

> *L'idée en question suggère que l'apparition des concepts chez les organismes capables de cognition est du même type que l'apprentissage du partitionnement des espaces d'activation neuronale*[56].

[56] Churchland Paul, *The engine of reason, the seat of the soul: a philosophical journey into brain.* First MIT Press paperback edition, 1996 (67, 1, 18-f)

Churchland n'affirme pas que les cerveaux sont des réseaux connexionnistes à trois couches, il considère plutôt ces réseaux comme des modèles approximatifs. Le problème, selon lui, est le rôle joué par la couche de sortie qui n'est là en fait que pour fournir au chercheur un moyen simple de contrôler la dynamique du réseau. Mais, il est prêt à accepter les deux premières couches comme lieu où s'opère la conceptualisation.

Est-ce vraiment le cas? Il nous semble que le système à lui seul n'a pas encore la capacité de construire un concept. Le travail de conceptualisation *bottom up* que tente d'expliquer ce modèle n'arrive à rien sans un travail *top down* opéré par un agent extérieur possédant déjà le concept. Rappelons que l'expérimentateur supervise l'apprentissage du réseau en imposant un vecteur de sortie cible. Sur ce point Churchland affirme :

> *Sur ces partitions et les neuf points de «visage familier» qu'elle comporte, les cellules de la couche 3 se sont lentement réglées avec succès.*

Ou encore :

> *Climbing up from the sensory periphery, we saw how feed-foward network can reconstruct various phenomena associated with sophisticated pattern recognition ... ... including the emergence, through repeated experience, of conceptual framework...*[57]

On voit l'idée d'un concept qui émergerait d'une mécanique autonome en amont, or il nous semble que l'on oublie vite ici que ce n'est pas les partitions de la couche cachée qui règle la couche de sortie, mais plutôt le vecteur de sortie souhaité du chercheur qui règle les deux couches synthétisantes et généralisantes. Par

[57] p. 211, f, f - 212, 1, 1

conséquent, son modèle seul ne peut pas produire de concepts, il doit obligatoirement tenir compte de l'humain derrière.

4-Réseau de neurones et hétérogénéité de Machery

Maintenant, à quel point ce type de modèle peut expliquer ou implémenter les trois types de concepts en psychologie? Selon plusieurs tels que McClelland (1994), dans un système biologique, le système de mémoire où logent les concepts se résume principalement à des populations de neurones situées dans le néocortex qui peuvent être simulées par les RN multicouches[58] tels que discutés par Churchland. Ces populations de neurones (ou aires dans le cortex) sont grandement interconnectées. De façon générale, ici aussi, une représentation est un patron d'activités distribuées à travers le réseau ou des parties du réseau. L'utilisation d'une représentation consiste à provoquer un patron d'activités par un autre patron qui sert de signal. Ces signaux peuvent être causés par exemple par des stimuli sensoriels ou par des associations inhérentes à l'apprentissage. Par exemple, les *patterns completion* sont des formes d'associations pour lesquelles un patron plus large, général, sert de signal pour un patron plus petit ayant comme contenu des épisodes spécifiques ou des informations sémantiques à propos d'un objet ou d'un évènement en particulier.

Cela dit, selon McClelland, il existe différents types de processus qui utilisent différents types de représentations: sémantique, encyclopédique et épisodique (9,4). Les types de représentations varient en fonction de leur utilisation dans des processus différents, de par leur contenu, mais structurellement ils ne sont pas distincts. Toutes les représentations sont entreposées dans le système néocortical suivant une accumulation de changements de connexions qui est le fruit de

[58] James L. McClelland, *Why there are Complementary Learning Systeme in the Hippocampus and Neocortex*, 1994

l'activation répétée de patrons d'activités superposés et de l'action de l'algorithme d'apprentissage à chaque répétition.

Si l'on tente de cibler les trois types de concepts en psychologie dans ce système, on s'aperçoit que le débat discuté au premier chapitre ne tient pas compte de cette architecture qui implémente les concepts. D'abord, physiquement les différentes représentations sont entreposées sur les mêmes populations de neurones. Ensuite, leur acquisition suit le même mécanisme avec de petits réglages différents.

Précisément, l'acquisition de représentations commence avec l'entreposage, dans des réseaux de neurones de l'hippocampe, de contenus à propos de choses spécifiques. Peut-on appeler cela des *exemples* ? Pas dans la mesure où cette information doit être intégrée ensuite par le néocortex avant d'être utilisée dans des processus de catégorisation, de raisonnement. Maintenant, ces représentations spécifiques ne se rendent jamais telles quelles dans le cortex. Le patron d'activité se dégrade dans l'hippocampe et faute d'autres présentations, l'apprentissage vers le cortex est trop lent pour mémoriser un exemple parfait. En effet, le cortex extrait seulement les éléments communs et généraux des patrons représentant des choses spécifiques dans l'hippocampe. Bref, l'architecture de la mémoire et de l'apprentissage nous laisse croire qu'il n'y a pas de vrais *exemples* (images fidèles) dans le cortex. Ces *exemples* sont déjà des représentations plus générales déjà en voie d'être des *prototypes*. En fait, la frontière entre un *prototype* et un (pseudo) *exemple* dans la mémoire corticale est peut-être une question de degré du taux de la règle d'apprentissage. En effet, la règle qui contrôle le taux du changement de poids de connexions qui suit chaque époque ou présentation à un réseau de neurones peut varier. Si cette règle a un taux élevé, le contenu à propos de choses spécifiques présentées au réseau provoquera un gros changement de connexion et par conséquent, il provoquera de l'interférence avec la structure déjà existante. À l'inverse si cette règle a un taux faible, l'apprentissage privilégiera la structure

représentationnelle en place et intègrera plutôt les grands aspects communs d'un contenu spécifique nécessaire à son utilisation future. Or selon McClelland, le cortex apprend très lentement, sinon notre mémoire se renouvèlerait à chaque *exemple* rencontré par notre cortex. Par contre, peut-être que le taux de cette règle varie très légèrement selon le contexte (ou l'information réinitialisée dans l'hippocampe) et qu'ainsi peut-être que parfois l'information est consolidée dans le cortex suivant un taux élevé et d'autres fois légèrement plus faible; peut-être que parfois l'information consolidée ressemble plus à un exemple et d'autres fois à un prototype. Mais il n'y a pas de distinction spéciale entre les deux si l'on tient compte de l'architecture qui les implémente.

En ce qui concerne les *théories,* elles ne sont pas étrangères à cette architecture. Dans la mesure où tous les patrons d'activités sont largement distribués à travers les mêmes populations de neurones et que chaque occurrence des patrons change la configuration des poids de connexions de ces populations, les représentations généralisantes sont superposées sur le réseau et par conséquent, elles partagent des représentations plus simples entre elles. Cette superposition imposée par l'architecture cognitive fait en sorte que toutes les représentations sont imbriquées dans un réseau sémantique. Par exemple, la représentation généralisante de *moineau* activera les propriétés *ailes, poumons, plumes, bec, yeux,* etc. Maintenant, la plupart de ces propriétés sont partagées par la représentation *oiseau* encore plus générale et d'autres propriétés sont partagées par des représentations encore plus généralisantes comme mammifères, *vivants, objets terrestres*, etc.

Bref, nous soutenons que la configuration des poids synaptiques d'une population de neurones remplit les fonctions des *théories*, alors que les fonctions des *exemples* (pseudo) et des *prototypes* sont remplies par les patrons d'activités acquis selon un taux légèrement variable de la règle d'apprentissage sur ce même réseau de neurones. Par conséquent, les trois types de représentations généralisantes

implicites/type1 que Machery aimerait diviser en trois espèces, constituent en fait trois caractéristiques d'un même complexe ou structure représentationnelle.

5– Critique de la conceptualisation internaliste telle qu'opérée par les RN

Les modèles connexionnistes fournissent un modèle de système internaliste qui semble pouvoir reconstruire de manière supervisée certains types de représentations généralisantès. Nous utilisons les termes «système internaliste» car ces modèles portent avec eux l'idée que tout le travail de conceptualisation doit être incarné sur un réseau de neurones et doit être produit par celui-ci. On le constate du fait que Churchland ou McClelland négligent toute explication *top-down* ou tout facteur situé à l'extérieur du RN. Or, la première étape de l'apprentissage d'un réseau de neurones est opérée par l'expérimentateur, en l'occurrence celui-ci détermine un type de sortie souhaitée en chargeant les cellules ou le comportement de sortie d'une signification. C'est-à-dire que l'expérimentateur associe une classe de choses (ou une représentation d'une classe) à détecter à une cellule de sortie ou un comportement, généralement prévue et discutée avant même de construire ou d'observer le réseau comme tel. Le travail de conceptualisation *bottom-up* que tente d'expliquer ce modèle n'arrive à rien sans un travail *top-down* opéré par un agent extérieur possédant déjà le *concept* (ce que Churchland nomme concept). En définitive, l'apprentissage demeure supervisé et orienté par l'expérimentateur et, par conséquent, on n'explique aucunement comment les concepts dans la tête de l'expérimentateur ont été créés à l'origine[59].

La question importante pour comprendre l'origine, depuis zéro, des concepts n'est pas seulement de savoir si un individu ou un robot (RN) peut apprendre ou être programmé pour interagir avec son environnement en utilisant des représentations

[59] Dans notre mémoire, nous étudierons des modèles plus récents qui apportent des réponses au problème de la supervision pour certains types de catégorisation.

ayant la fonction de concept. Mais plutôt, est-ce qu'un individu ou un robot peut de manière autonome créer un concept, une structure sémantique qui pourra jouer le rôle des représentations conceptuelles?[60]

L'autre grand problème est que même si le partitionnement des vecteurs est une étape de base essentielle à la conceptualisation, elle n'explique pas encore tout le phénomène de conceptualisation selon nous. Il faudra expliquer quel est le rôle de la classification, de la généralisation et de la stabilisation de la perception dans la cognition globale et particulièrement dans le travail linguistique et réfléchi. Car si on se borne à ces derniers critères *implicites* pour définir un concept, alors on devra peut-être considérer une machine distributrice de coca comme un agent cognitif qui conceptualise, puisqu'elle peut classer les argents, les canettes et peut même généraliser en acceptant de l'argent qu'elle n'a jamais vu. En fait, le travail de catégorisation doit aussi répondre aux autres critères explicites (de type 2) avant de parler de concept.

Nous traiterons dans ce chapitre et le suivant que du premier problème, à savoir du problème de la supervision et de ses fondements internalistes.

5.1-Problème de la supervision de l'apprentissage

Churchland n'est pas vraiment ignorant du problème de la supervision de l'apprentissage. Il est seulement secondaire à ses yeux. Il dit rapidement (p.270)[61] :

> *The vocabulary already in place and already at work in the local cognitive commerce forms an abstract template that shapes the infant brain's development by narrowing its search space during learning.*

---

[60] Luc steel, *The symbol grounding problem has been solved so what's next,* 2008

[61] Churchland Paul, *The engine of reason, the seat of the soul: a philosophical journey into brain*. First MIT Press paperback edition, 1996

Toutefois, il ne décrit jamais le commerce cognitif entre les agents. Il n'explique pas comment l'espace abstrait qui sert d'orienteur aux apprenants est construit ou est transmis réellement. Pour Churchland, les concepts et l'espace culturel sont donnés et le réseau de neurones doit traiter ces données.

Churchland, donc, ne cherche pas à décrire l'origine des concepts, depuis zéro, ni leur existence dans un commerce social, il prend l'individu cognitif comme unité de base et cherche à décrire les mécanismes internes de ce dernier sans se soucier de l'interaction avec quoi que ce soit d'autre.

Pour nous, qui cherchons à définir la réalisation de concepts, l'explication de la façon dont un individu procède au traitement des informations données n'est pas suffisante. Nous ne cherchons pas à décrire les réseaux de neurones comme tels, mais nous cherchons dans les descriptions des réseaux de neurones une explication à la création de concepts. Si la lunette centrée sur l'individu comme unité de base ne nous explique pas tout, il faut regarder ce qu'une autre plus large pourrait nous apporter.

6-Critique de la conceptualisation internaliste en général

6.1-Éclaircissement du concept de «construction»

Les RN ne réussissent pas à expliquer la création des concepts non pas parce que la technologie ou le savoir est trop jeune, mais bien parce qu'ils sont cloisonnés dans une perspective internaliste. Ce problème s'étend à plusieurs auteurs en sciences cognitives qui négligent ce fondement théorique. Voyons plus précisément comment cette impasse se dessine en général et pourquoi elle demeure ignorée par plusieurs.

Nous avons montré dans le premier chapitre qu'une large partie des sciences cognitives chapeautées par la psychologie a su faire renaître le concept de concept en l'harmonisant à son objet d'étude (l'individu), à ses méthodes (plus empiriques) et à ses thèses en général. Il est devenu désormais une entité psychologique ou interne décrite avec un vocabulaire propre à cette discipline. Par le fait même, le concept de concept s'est différencié des descriptions et des explications philosophiques, si bien que certains tels que Machery discutent aujourd'hui de deux choses incomparables[62]. En effet, selon lui, en sciences cognitives, un concept est une sorte de représentation mentale stockée dans la mémoire à long terme, alors qu'en philosophie, un concept à propos d'une chose x est ce qui est nécessaire et suffisant de posséder pour avoir des attitudes propositionnelles de x comme un x (x as x).

Or, dans la littérature en sciences cognitives, la philosophie et la psychologie se côtoient toujours, ce qui provoque parfois certaines confusions concernant l'utilisation du concept de concept. Si l'on ajoute à cet imbroglio psycho-philosophique le fait que plusieurs autres disciplines associées aux sciences cognitives s'intéressent aux concepts, telles que l'anthropologie, l'informatique ou la sociologie, il n'est pas surprenant d'observer des auteurs utiliser les mêmes termes pour discuter de choses différentes.

En fait, certaines confusions terminologiques sont inhérentes à l'objet général qu'étudie chaque discipline. Par exemple, on l'a vu, une «inférence» en psychologie est essentiellement un processus mental qui se situe dans un individu (puisque son objet général d'étude est ce dernier), alors que pour un mathématicien une «inférence» peut exister en dehors des limites de l'individu. Ainsi, lorsqu'on discute par exemple de la production d'une inférence, pour le psychologue cela réfère aux mécanismes cognitifs propres à la manipulation de certaines informations dans un

[62] Machery, *Doing without concept,*

individu, alors que pour le mathématicien, suivre mentalement une inférence peut ne rien avoir à faire avec la nature de cette dernière.

Ainsi, lorsqu'on tente d'expliquer *la création ou la construction des concepts depuis leur origine,* plusieurs présupposés peuvent provoquer de la confusion. En effet, en sciences cognitives, une grande partie de la littérature sur les concepts adopte le point de vue internaliste, tel qu'on l'a vu avec la psychologie dans le premier chapitre ou avec les RN dans ce chapitre-ci, selon lequel toutes explications concernant la *construction* des concepts réfèrent à des entités, des phénomènes ou des processus qui se situent dans les limites de l'individu. Par conséquent, la question de l'origine d'un concept est souvent confondue avec celle qui consiste à savoir comment un concept *arrive* dans un individu. Dans ce sens, on discute davantage de l'acquisition, de l'apprentissage et de la *construction* d'un concept par un individu que de l'invention, de la création ou de la construction d'un concept en soi. Par exemple, on tente moins d'expliquer dans la perspective internaliste comment le concept «table» a été construit (dans le sens de créé) à son origine et transmis par la suite, que d'expliquer comment un enfant l'acquiert ou le (re)construit *dans sa tête*. En d'autres termes, une théorie psychologiste et internaliste n'a besoin que d'une définition des concepts limitée aux intérêts psychologistes: expliquer les processus mentaux.

Or, nous croyons qu'il y a une certaine ambiguïté dans l'utilisation de l'expression *construction de concepts* dans la littérature propre aux sciences cognitives qui participe au fait que l'on croit expliquer toute l'étendue de la création d'un concept, alors que l'on explique souvent que sa reconstruction ou sa transmission. Éclaircissons les termes.

D'abord, dans le paradigme internaliste, *l'acquisition de concepts* réfère aux processus cognitifs qui permettent la formation d'un concept dans un individu; concept qui existe généralement déjà ailleurs, dans d'autres individus, voir dans leur

mémoire à long terme ou qui est donné par le langage. Ensuite, *l'apprentissage d'un concept* réfère à un mode *d'acquisition* par un individu qui est supervisé par un expérimentateur ou un enseignant. L'expérience typique où l'on observe *l'apprentissage* d'un concept par un individu implique qu'un autre individu connaissant fournisse de la rétroaction au sujet durant son apprentissage.

Finalement, la *construction d'un concept* réfère dans la littérature internaliste à un mode *d'acquisition* non supervisé[63], contrairement à l'*apprentissage*. Ici un individu peut se construire un nouveau concept pour catégoriser certains objets qu'il observe, mais qui n'entreraient pas dans un concept qu'il possède déjà. Toutefois, si l'on se fie aux expériences en laboratoire décrites dans la littérature, la construction d'un nouveau concept ne semble nouvelle que pour le sujet de l'expérience: l'expérimentateur connaît le concept. En effet, lors d'une expérience type selon Murphy, un expérimentateur expose à un sujet des choses qu'il considère déjà comme étant membres d'un concept (ou l'inverse, Quinn 87) en s'attendant à ce que le sujet reconnaisse seul que ce sont le même genre de choses. Généralement, à partir de concepts très bien définissables, les sujets réussissent à (re)construire des concepts sans *feedback* de l'expérimentateur. Bref, ici le concept existe déjà dans la tête de l'expérimentateur qui construit l'expérience en fonction de celui-ci. De plus, l'observation de la construction d'un concept se fait le plus souvent chez les jeunes enfants qui ne possèdent pas encore de concepts pour la majorité des choses. Entendons ici qu'il semble difficile de rassembler un groupe d'items dont aucun adulte n'aurait de concepts déjà en mémoire pour le catégoriser, mais dont la construction d'un nouveau concept serait facile pour cet adulte qui entrerait en contact avec ce groupe d'items. C'est-à-dire que les adultes en général possèdent un très grand nombre de concepts pour catégoriser les choses nouvelles et que rarement ils en construisent, seuls, de nouveaux. Dans ce cas, les termes *construction d'un concept* semblent signifier une reconstruction sans *feedback* dans un individu enfant

[63] Murphy, ...

d'un concept qui existe déjà dans la tête de l'expérimentateur ou des adultes en général.

Certains pourraient nous objecter ici que ce type d'expérience reproduit ce qui pourrait se passer chez un individu seul, sans expérimentateur, face à des instances d'une espèce naturelle. Peut-être, mais pourquoi ne pas discuter de ce que reproduit (ou ce à quoi correspond dans la réalité) le rôle du chercheur? L'environnement défini par ce dernier a bien une fonction essentielle non explicitée dans ce type d'expérience, alors qu'est-ce qui remplit cette fonction dans la *réalité*?

Ici il faut distinguer, au minimum, le cas où un individu se retrouverait pour la première fois devant une instance déjà conceptualisée par d'autres, comme dans le cas des simulations en laboratoire, et le cas où un individu ferait face à une chose inconnue de tout humain. Dans le premier cas, même dans la réalité, la construction d'un *nouveau* concept ne semblerait nouvelle que pour le sujet. Un individu qui se retrouverait pour la première fois devant un *cheval de trait*, par exemple, ne pourrait pas conceptualiser cette instance à partir de zéro à l'aide de ses seuls mécanismes de catégorisation. Il utiliserait les centaines de concepts qui existaient avant sa naissance et qu'on lui a transmis. En effet, des populations sur des générations d'individus ont préparé conceptuellement ou culturellement son expérience individuelle. Il partirait au minimum des concepts «vivant», «animal», «ferme», etc., que son entourage immédiat lui a permis d'acquérir. Un *enfant sauvage* n'y arriverait sûrement pas seul, avec ses mécanismes innés. Par conséquent, les termes *construction de concept* signifieraient encore (comme en laboratoire) une reconstruction dans un individu d'un concept. Et la question demeure ouverte: si un concept existe déjà avant la construction de celui-ci dans un individu, d'où vient-il à l'origine? L'expérience en laboratoire ne répond pas à cette question.

Maintenant, même dans le deuxième cas qui dépasse ce que peuvent reproduire les expériences en laboratoire, c'est-à-dire si un individu se retrouvait pour la première fois devant une chose totalement inconnue de tous, qui n'est pas subsumée sous un concept, il ne pourrait pas conceptualiser cette instance d'espèce mystérieuse à partir de zéro. Il partirait encore de concepts existant ou bien il risquerait de ne pas la saisir du tout. La plupart des maladies étaient inconnues il y a 3000 ans malgré la myriade de symptômes que l'on pouvait observer. Picasso n'aurait pas pu développer le cubisme au 12^{e} siècle comme Einstein n'aurait pas pu conceptualiser la relativité du temps et de l'espace à l'époque d'Homère. La construction, depuis zéro, des concepts ne semble pas se faire à partir d'un seul individu.

Par conséquent, il est légitime de se demander encore d'où vient le concept dans la tête de l'expérimentateur. L'a-t-il construit seul ou l'a-t-il appris?

6.2-Conclusion

Traditionnellement, ce type d'impasse est vite solutionné par l'invocation de processus ou d'entités représentationnelles innées. Prenons en exemple les questions suivantes : pourquoi plusieurs personnes peuvent construire les mêmes concepts? Pourquoi l'enfant construit *le* concept attendu par l'expérimentateur? Si l'on cherche avec des présupposés psychologistes ou internalistes, l'on risque de chercher les causes dans l'*individu.* Par exemple, en psychologie, les concepts sont compris comme des espèces naturelles (exemple, prototype, théorie, idéal, etc.) qui possèdent des mécanismes d'acquisition universels propres à chacun d'eux; c'est pourquoi différents individus peuvent partager des contenus semblables. En d'autres termes, les concepts ont la structure qu'ils possèdent chez tous les individus à cause de leur faculté innée de les acquérir et de les utiliser; la question à savoir pourquoi l'enfant reconstruit le même concept que l'expérimentateur est relayé à un niveau d'explication biologique.

Par contre, nous pouvons toujours répliquer qu'on oublie trop souvent le rôle de l'environnement dans lequel l'individu est nécessairement plongé. Selon les positions internalistes-innéistes orthodoxes[64], son rôle serait seulement de fournir les stimuli nécessaires pour encadrer le développement de la faculté d'acquérir, comme la lumière ou la nourriture permettent au système visuel et moteur de se développer.

Toutefois, le bagage inné d'un individu ne semble pas constituer la seule source d'explication de l'origine des concepts, pour plusieurs raisons; nous en avons soulevé au moins deux. D'abord, tant le type d'expérience décrite plus haut avec les RN que dans celles décrites par Murphy, il ne faut pas oublier que la première étape de la conceptualisation est opérée par l'expérimentateur, en l'occurrence celui-ci détermine un concept qu'il souhaite voir construire par l'enfant ou le RN. Le travail de construction *autonome* que tente d'observer l'expérimentateur n'arrive à rien sans un travail de supervision et d'orientation opéré par un agent extérieur possédant déjà le concept: l'expérimentateur propose des choses qu'il a choisies parce qu'elles étaient membres d'un même concept. Par conséquent, on n'explique pas encore comment le concept, dans la tête de l'expérimentateur, a été construit à l'origine[65].

De plus, il semble évident, même en psychologie, qu'un individu isolé ne puisse pas reconstruire tous les concepts qui existent déjà. Personne ne peut s'éduquer lui-même en confrontant ses mécanismes représentationnels à son environnement. Il est préférable d'enseigner à un enfant des années durant des versions officielles du français, des mathématiques, de l'histoire, de l'informatique, etc. L'apprentissage semble, dans cette mesure, un mode d'acquisition peut-être plus important que la

---

[64] Kirby, S., Dowman, M. and Griffiths, T. (2007). *Innateness and culture in the evolution of language*. Proceedings of the National Academy of Sciences.

[65] Il se pourrait que des modèles de l'évolution par algorithme génétique de concepts contredisent en partie cette critique. Cette possibilité pourrait être étudiée et corrigée le cas échéant dans un prochain travail.

construction. Alors d'où viennent les concepts appris dans ces matières si leur origine ne se résume pas aux mécanismes cognitifs individuels? Comment expliquer le fait que l'acquisition est supervisée et orientée par des concepts déjà existants, tout en soutenant que les concepts sont déterminés par ces mécanismes d'acquisition? Qu'est-ce qui est venu en premier, les concepts ou les mécanismes d'acquisition?

Nous défendons la thèse qu'il y a quelque chose de plus à l'extérieur de l'individu qui participe aussi à la construction des concepts depuis leur origine. En nous appuyant sur des auteurs de différentes disciplines tels que Hazlehurst, Hutchins, Kirby, Donald, et Dennett, nous soulevons l'hypothèse que, si l'on considère le rôle de *la transmission culturelle* pour décrire les concepts, on s'aperçoit que les régularités de ces derniers ne reposent pas tant sur des contraintes internes et innées fortes, et surtout, on s'aperçoit que les concepts peuvent être accumulés et enrichis dans le temps et à travers les générations dans une *structure interindividuelle*. En ajoutant cette structure sociale à celle individuelle, on prend une communauté d'esprits comme unité d'analyse au lieu de prendre l'individu isolément. Ainsi, depuis cette perspective, on peut soutenir que des populations peuvent construire des concepts impossibles à construire pour des individus. Finalement, on peut expliquer comment peut émerger un concept dans une population d'individus sans que personne de l'extérieur ne le donne.

C'est l'une des pistes de solutions que nous aborderons dans le chapitre 4. Nous y étudierons la possibilité de créer des concepts, depuis zéro, à partir d'une population de RN classiques sur plusieurs générations. Mais, auparavant, dans le prochain chapitre, nous nous distançons d'un pas philosophique pour essayer de comprendre encore davantage l'origine historique même du paradigme internaliste.

CHAPITRE III

INTERNALISME COMME INDIVIDUALISME

Comment les concepts sont-ils créés depuis leur origine? À cette question, nous avons proposé à la fin du dernier chapitre que l'on peut reconnaitre les réponses générées par le courant dominant en sciences cognitives en ce qu'elles se rapportent à une conception internaliste. De façon générale, ce paradigme regroupe un ensemble d'explications des processus ou mécanismes cognitifs et des représentations qu'ils traitent en focalisant sur ce qui se trouve dans une individualité. L'intelligence artificielle s'intéresse aux mécanismes dans un robot; la psychologie, aux processus cognitifs dans un sujet humain; les neurosciences, aux mécanismes neurologiques réalisés par un système nerveux. Conséquemment, les théories sur les concepts en sciences cognitives qui s'inscrivent dans ce courant dominant partent souvent de l'idée qu'il faut expliquer comment un individu traite des concepts et non de l'idée de décrire les concepts en soi. Sans nier les thèses internalistes, nous pensons que ce seul point de vue n'explique pas totalement la nature des concepts. En effet, un individu ne peut pas construire à lui seul, à partir de zéro, l'ensemble des concepts qu'il a en mémoire.

1-Internalisme en sciences cognitives

Au premier chapitre, nous proposions une définition du concept de concept qui est grandement influencée par les théories psychologiques. En effet, un nouvel engouement pour les concepts est apparu en partie en psychologie au 20^{e} siècle et plus généralement dans les sciences cognitives où l'on a redéfini trois principales familles de théories: les théories des *exemples*, des *prototypes* et des *théories*[66].

De plus, les différentes disciplines reliées aux sciences cognitives ont porté leur regard plus empiriste sur la nature des concepts ce qui a entrainé la nécessité, nous l'avons vu, que les propriétés attribuées aux concepts s'accommodent de la façon qu'ont les individus d'utiliser et d'acquérir des concepts lors de tâches expérimentales.

Dans le même sens, selon Machery, l'on soutient dans la littérature que les gens catégorisent et raisonnent de la façon qu'ils le font puisque ces processus utilisent des concepts qui possèdent toutes les mêmes propriétés. C'est-à-dire que la façon dont les concepts arrivent dans la mémoire (mécanisme de catégorisation par exemple) et sont utilisés par la suite (déduction par exemple) est relative à la nature des concepts. En l'occurrence, les trois principales théories sur les concepts en sciences cognitives proposent différentes explications de jugements catégoriels et de raisonnements (inductifs ou déductifs) qui correspondent aux différentes définitions des concepts (et/ou l'inverse). Par exemple, parfois un individu catégorise en comparant une représentation X avec une somme statistique de propriétés typiques d'un groupe de représentations mémorisées (théorie des prototypes). Parfois, un individu catégorise en comparant une représentation X avec des membres

[66] Selon Piccinini : Voir Hampton [1993] pour les prototypes, Nosofsky [1988] pour les exemples, et Gopnik and Meltzoff [1997] pour les théories

importants d'un groupe de représentations mémorisées (théorie des exemples). Parfois, il catégorise si une représentation X a les propriétés qu'il s'attend à voir chez les membres d'un groupe de représentations en vertu des causes qui font que les membres sont ce qu'ils sont (théorie des théories).

Bref, la nature des processus mentaux propre à la cognition supérieure est relative à la partie fonctionnelle de la nature des concepts et réciproquement[67]. Or, un des aspects essentiels des processus mentaux en général selon le courant dominant en sciences cognitives est qu'ils sont identiques à des processus cérébraux, ou réalisés exclusivement par ceux-ci.

Selon Mark Rowland[68] par exemple, les sciences cognitives se regroupent dès leur origine (années 60) autour du postulat affirmant que les processus cognitifs sont des «programmes» abstraits réalisés dans le «hardware» du cerveau. Cette analogie avec l'informatique qui permet la réunion de la psychologie (programme) et des neurosciences (hardware) sert de fondement théorique sur lequel se développe, depuis, l'étude scientifique des processus cognitifs. Et ce, même si cette dernière se développe de plusieurs manières différentes. En effet, au départ, sous l'impulsion des réussites en intelligence artificielle, l'emphase est davantage mise sur les aspects abstraits et fonctionnels («programme») de la cognition. Plus tard, dans les années 80-90, l'emphase se déplace davantage sur les aspects physiques («hardware») avec les approches connexionnistes que l'on a décrites aux chapitre 2. Mais, peu importe si l'emphase est mise davantage sur les aspects fonctionnels ou physiques ou les deux, la trame commune de ces études se résume à l'hypothèse selon laquelle les processus cognitifs se produisent toujours à l'intérieur de têtes d'organismes

[67] Cela est moins vrai en ce qui concerne le contenu, ce que nous allons voir plus loin.
[68] Rowland, Mark, *The new science of the mind,* The MIT Press, 2010

pensants, ce qu'on a soulevé comme critique au chapitre précédent et que Rowland nomme la *Science cognitive cartésienne* (Cartesian cognitive science)[69].

Partant, si les processus mentaux tels que la mémorisation, la catégorisation et le raisonnement sont ultimement identiques à des opérations cérébrales, alors le concept de concept entretient le même type de relation d'interdépendance avec les opérations cérébrales qu'avec les processus mentaux. C'est-à-dire, que si d'une part la nature des concepts est relative à la nature des processus mentaux (relation de réciprocité), et que d'autre part les processus mentaux sont identiques aux processus cérébraux, alors la nature des concepts est relative à celle des processus cérébraux propre à la cognition supérieure (relation de réciprocité). Nous allons nous référer à cette thèse ou a priori théorique en tant qu'*internalisme* ou *paradigme internaliste en science cognitive*.

Problème avec l'internalisme. La difficulté avec ce paradigme n'est pas qu'il repose sur une explication matérialiste ou biologique. Le problème est plutôt qu'en adoptant le *cerveau d'un individu* comme seul objet d'étude, comme seule source des concepts, la définition de la nature des concepts est réduite à la façon dont *un individu* procède à leurs traitements. C'est-à-dire que la description et l'explication des concepts sont fonction des prérogatives de l'étude scientifique de la cognition *d'un individu.* Cette réduction de la nature du concept à celle de l'individu laisse en plan plusieurs questions ou problèmes dont certains ont été soulevés au chapitre 2. La principale question, selon nous, est celle de l'origine des concepts, incluant leurs contenus. Cette question, nous le verrons, suscite à son tour la question de la normativité et du caractère public des concepts.

[69] Rowland, Mark, *The new science of the mind,* The MIT Press, 2010

Dans l'immédiat, la question qui nous intéresse est: Qu'est-ce qui motive plusieurs auteurs représentatifs du courant dominant en sciences cognitives à continuer de réduire ou de cloisonner la description et l'explication des concepts à celles de leur acquisition par *un* cerveau? Même si cela implique de traiter avec indifférence ou carrément d'omettre la question de l'origine ultime des concepts.

2- Problème d'attribution de la source des concepts

Pour bien saisir le problème, il est pertinent de prendre un recul théorique et d'essayer d'appréhender le contexte plus large ou l'idéologie en arrière-plan dans laquelle s'inscrit ce recours tacite à un individu et l'intérieur de son crâne comme seule source des concepts.

On trouve un courant secondaire dans la littérature où certains auteurs (W.Prinz, Dennett, Wilson, Michon) partagent l'idée que le paradigme internaliste est engendré et maintenu par des tendances socio-historico-culturelles. En d'autres termes, qu'il est un construit social créé et préservé par un paradigme beaucoup plus large, soit par un individualisme comme idéologie qui sert de trame à toute la culture occidentale.

Pour comprendre cette hypothèse, il faut aussi d'une part faire le lien entre l'idée reçue soutenant que les représentations généralisantes sont générées et exécutées par *un organe cognitif* (organ of mind) et l'idée plus ancienne stipulant que l'humain possède un espace psychologique interne qui est la source de ces représentations, mais aussi des pensées et de l'action en général. Il faut montrer que la première idée, l'organe cognitif, est hérité de la deuxième, l'idée de l'espace interne personnel, à savoir, que l'internalisme en science cognitive est une entreprise de naturalisation

de l'espace psychologique interne personnel, le «soi» mental [70]. Ensuite, il faut expliquer que ce concept de «soi» mental auquel l'on attribue généralement la source des concepts est en fait un construit théorique grandement utile, voire nécessaire, pour comprendre tant les caractéristiques politiques, économiques, épistémologiques, sociologiques, éthiques, que religieuses de notre culture. Bref, il faut montrer que cet espace interne personnel est une fiction hautement utile pour les sociétés individualistes, tellement, que les savants tentent d'en faire *une espèce naturelle.*

La pertinence de cet argument historique tient de l'idée que si l'internalisme et particulièrement l'attribution de la source des concepts à un individu est un construit social, alors cette entreprise que poursuit les sciences cognitives est contextuelle, relative et susceptible de ne pas correspondre adéquatement à la réalité. En d'autres termes, le programme de naturalisation de la conceptualisation qui consiste à fonder cette dernière *essentiellement* sur les mécanismes cérébraux est relatif à un certain type d'organisation sociale. De plus, si l'exigence d'une ontologie naturaliste et matérialiste est justifiée et nécessaire selon nous, le projet de vouloir à tout prix trouver ou faire exister des construits socio-culturels dans des cerveaux est moins justifié. Partant, cet argument ouvre la voie toute grande à notre thèse, à savoir que l'internalisme n'arrive pas, seul, à expliquer l'origine des contenus des concepts adéquatement.

Nous verrons d'abord le lien entre l'arrière-plan individualiste, l'espace interne de l'individu et les concepts. Ensuite, nous verrons comment cet espace interne est apparu dans l'histoire humaine précisément pour résoudre un problème d'attribution de la source des représentations en général.

---

[70] W.Prinz, *Emerging selves: Representational foundations of subjectivity,* www.elsevier.com/locate/concog, 2003

2.1-Paradigme individualiste/ paradigme internaliste/ concept de concept

L'individualisme comme concept implique d'une part l'idée d'une certaine relation entre l'individu et la collectivité, et d'autre part, il sous-tend une conception de la nature humaine.

D'abord, *du point de vue social*, l'individu humain y est compris comme relativement autosuffisant et libre; il n'est pas pensé comme un être déterminé par un groupe duquel il ne serait qu'une partie subordonnée. Une société qui suit une idéologie *(un système d'idées et de valeurs qui a cours dans un milieu social donné*[71]*)* individualiste possède une organisation qui permet et protège cette liberté et cette autonomie. Par exemple, la liberté de posséder des biens et de commercer, la liberté de conscience, d'expression et d'association, la liberté en matière de mœurs, aussi la libre mobilité géographique et professionnelle, la protection des droits individuels[72] sont toutes des caractéristiques d'une société individualiste. Plus une structure sociale en possède, plus elle vise l'indépendance des individus.

Ensuite, concernant *la conception de l'humain*, disons d'emblée que dans une idéologie individualiste, la liberté individuelle est avant tout une liberté de penser (cognitive) et d'action. En effet, une action ou une pensée relève d'un individu, il en est responsable, dans la mesure où l'on présume qu'il peut la penser seul, la choisir librement et la commettre lui-même si c'est une action. À l'inverse, si un comportement ou une pensée n'est pas commis librement, mais sous une influence *extérieure* ou inconsciente, il ne sera pas imputé à l'individu directement[73]. Par exemple, la justice canadienne reconnait qu'un meurtrier est pleinement responsable s'il a prémédité (librement) son action, moins s'il a agi impulsivement et encore

[71] Louis Dumont, *Essais sur l'individualisme,* Seuil, 1983
[72] Alain Laurent, *Histoire de l'individualisme*, Presse universitaire de France, 1993.
[73] Kant, Fondement de la métaphysique des mœurs

moins si l'action est *survenue alors qu'[il était atteint] de troubles mentaux qui [le] rendaient incapable de juger de la nature et de la qualité de l'acte ou de l'omission*[74]. De la sorte, il semble que la liberté qui définit l'individu soit reliée en partie à la capacité de penser et, ici, de penser l'action. C'est-à-dire que l'individualisme implique une conception de l'homme et de ses propriétés cognitives qui en font un être libre, autonome et responsable.

Chez les philosophes phares de l'individualisme, l'on attribue à l'âme, à la raison (et ses facultés: la volonté par exemple) et/ou à la conscience cette capacité de penser et d'agir. Par exemple, selon Kant, la raison est libre pour autant qu'elle agit *selon une loi qu'elle s'est donnée à elle-même; pour autant qu'elle se considère elle-même comme l'auteur de ses principes*[75] qui guident l'action. Cette possibilité de penser ou de se donner à soi-même un objectif de manière autonome est nécessaire pour comprendre la notion d'individu. S'il ne possédait pas cette liberté, il ne serait pas lui-même la cause de ses pensées ou de ses actions et tout l'individualisme comme idéologie serait sans fondement[76].

De plus, pour les mêmes raisons, cette capacité de penser et de choisir préalablement son action doit se faire à l'intérieur de l'individu. Si elle se fait à l'extérieur, alors l'individu ne sera plus autonome, mais déterminé par quelque chose d'étranger à lui-même. Maintenant, que l'on nomme cette faculté interne raison, volonté, pensée, conscience, intentionnalité, imagination, elle constitue un espace interne virtuel où sont traités de manière idiosyncrasique des pensées et des scénarios d'actions potentiels qui rendent possible le libre arbitre. C'est-à-dire, que l'humain compris comme libre se représente en lui-même les actions qu'il peut, qu'il veut et qu'il va faire avant de les poser dans le monde[77]. Or, au meilleur de nos connaissances

---

[74] Article *16. (1) du Code criminel canadien, http://laws-lois.justice.gc.ca/fra/lois/C-46/*
[75] Kant, Fondement de la métaphysique des mœurs
[76] Par exemple : Kant, *Métaphysique des mœurs*
[77] Dennett, D. C., *Freedom evolves*, Penguin Book, 2004

aujourd'hui, quel type de représentations est utilisé pour générer des pensées et des scénarios d'actions? Les concepts. En effet, si l'on revient un instant à la définition des concepts en sciences cognitives, on se souvient que ces derniers sont *les ensembles de base de la connaissance (bodies of knowledge) communes à toutes les opérations de la cognition supérieure*[78]. Partant, ils doivent participer aussi à générer des pensées et tester des scénarios virtuels d'actions. Sinon, de quoi sont faites ces pensées?

Si c'est le cas, les concepts deviennent essentiels pour décrire et expliquer le type d'humain que suppose les sociétés individualistes, c'est-à-dire des humains qui gèrent seuls leur liberté, leur autonomie, leurs responsabilités à partir de leur for intérieur[79].

Bref, le raisonnement est le suivant : le lien entre le paradigme individualiste au sens large et le paradigme internaliste est précisément le fait que cet espace virtuel (objet d'étude de l'internalisme) est essentiel pour soutenir le concept d'individu (objet de l'individualisme) en ce que le premier est le lieu où s'exerce ce qui définit le deuxième: le libre-arbitre (ou quelque chose de la sorte). Et puisque les concepts sont les entités de base de tout ce qui se trame dans cet espace interne personnel, alors ils sont eux-mêmes des outils théoriques importants pour toute l'idéologie individualiste. Ainsi, un membre d'une société individualiste, lorsqu'il tente d'expliquer la nature des comportements humains doit référer à cet espace virtuel interne implicite et décrire les phénomènes mentaux et leurs contenus, quelque chose comme les concepts, nécessaires aux actions individuelles.

---

[78] Machery, *Concepts Are Not a Natural Kind*, Depart. of History and Philosophy of Science, University of Pittsburgh. 200?

[79] Dumont, Louis, *Essais sur l'individualisme*, Seuil, 1983

2.2-Fondation représentationnelle de la subjectivité

Pourquoi ce détour théorique est-il pertinent? Précisément parce que cet espace interne autonome et personnel, le «soi» mental, compris comme source des pensées, des actions et partant des concepts est en fait un construit socioculturel maintenu et préservé par les institutions linguistique, politique, scientifique, juridique, etc. Par conséquent, les propriétés (source libre et autonome des pensées) attribuées à cet espace interne ne sont pas constitutives d'un organe biologique de la pensée; leur explication ne se résume pas à une description des mécanismes cognitifs internes. Notez que nous ne disons pas que le fait d'attribuer la source de la pensée (et de la cognition) en général à un individu et son cerveau est une pure fiction (nous demeurons strictement matérialiste). Mais attribuer l'origine du *contenu* des pensées, et partant des concepts, à des individus autonomes et leur cerveau est une fiction utile et maintenue par nos institutions[80].

W.Prinz justifie cette thèse à l'aide de trois arguments. Primo, il montre comment l'émergence de la conscience et de ses contenus est liée à celle du «soi». Deuzio, il soutient que le «soi» mental est apparu pour solutionner un problème d'attribution de la source des représentations. Et finalement, il montre que certains mécanismes d'interactions sociales et certains discours maintiennent et préservent l'institution du «soi» mental.

Nous exposerons ces trois arguments et montrerons ensuite comment ce construit est à l'origine du paradigme internaliste en science cognitive.

---

[80] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

2.2.1-Expérience consciente et subjectivité

Contrairement aux roches ou aux moules, par exemple, lorsqu'on discute d'animaux proches des humains et surtout des humains comme tels, il semble pour plusieurs que les sciences naturelles ne peuvent pas tout expliquer. Nous attribuons à ces êtres des espaces internes qui ressemblent en gros à notre expérience consciente personnelle, expérience qui aurait un caractère subjectif[81].

Quelle est la nature des contenus de pensées conscientes que l'on situe dans cet espace interne? Car, il peut y avoir des contenus conscients ou inconscients nous dit W.Prinz. On n'a qu'à penser aux tests d'attention de Simons et Chabris[82] (le célèbre «moon walking bear»[83]) qui démontrent bien que certains contenus perceptuels n'accèdent pas à la conscience. Ces derniers contenus ne feront pas partie du vécu personnel dans la mesure où ils ne seront pas considérés comme des contenus constituants de l'expérience subjective de l'individu, du «soi». À l'inverse, les contenus mentaux conscients sont caractérisés par ce caractère subjectif, personnel, à savoir, par la présence implicite du «soi» mental. En d'autres termes, les représentations conscientes d'une situation se terminent précisément lorsque le «soi», l'expérience personnelle, s'en écarte[84]: sinon comment expliquer le changement de vitesse que l'on effectue inconsciemment alors que nous sommes occupés à une discussion?

Maintenant, pour montrer que l'espace interne personnel est en partie un construit socio-culturel, il faut expliquer comment la nature consciente des contenus mentaux est apparue. Or, puisque la relation entre les contenus mentaux et la présence

---

[81] Ibid.
[82] Simons et Chabris'
[83] Voir par exemple : http://www.youtube.com/watch?v=Ahg6qcgoay4
[84] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

implicite d'un «soi» mental caractérise leur nature consciente, il faut alors expliquer la constitution du «soi» mental et de ses représentations implicites pour expliquer la constitution de la nature consciente des contenus. Il faut expliquer à quoi sert le «soi» mental et pourquoi il a émergé durant l'évolution. Et pour W.Prinz, cela revient à expliquer comment nous sommes passés de bêtes sans «soi» à des bêtes qui en possèdent un.

Toujours selon W.Prinz, dans la littérature en science cognitive (Dennett, Donald, Edelman, Jaynes, Metzinger) plusieurs points communs sont partagés pour expliquer le développement du «soi» : (1) il est une évolution proprement humaine, (2) le «soi» mental introduit des avantages cognitifs et dynamiques et (3) il y a des conditions sociales et politiques qui permettent son apparition.

2.2.2- L'émergence des «soi»

Pour expliquer le passage de la bête sans «soi» à la bête avec un «soi», Prinz reprend la différence entre la créature popérienne et celle grégorienne de Dennett[85]. Ce dernier nomme les créatures poppériennes en l'honneur de Karl Popper qui aimait à dire qu'en «mettant les théories en compétion, cela permet à nos hypothèses de mourir à notre place»[86]. Pour Dennett, les créatures poppériennes ne mettent pas des théories sous forme de langage en compétition, mais des simulations de comportements possibles dans l'environnement. Ces simulations sont opérées dans un espace interne, mais qui ne constitue pas tout à fait encore un «soi». En résumé, cet espace interne sert à simuler les comportements possibles de telle manière que les actions non pertinentes dans l'immédiat soient écartées avant d'être expérimentées réellement. Toutefois, si ce n'est pas seulement les humains qui

---

[85] Voir par exemple : Dennett, D. C., *The Self as a Center of Narrative Gravity, in F. Kessel, P. Cole and D. Johnson, eds, Self and Consciousness: Multiple Perspectives, Hillsdale, NJ: Erlbaum, 1992. Danish translation, "Selvet som fortællingens tyngdepunkt," Philosophia, 15, 275-88, 19861986*

[86] Ibid.

possèdent ce genre d'habiletés virtuelles, il reste que les créatures poppériennes n'ont pas encore la moindre idée de ce qu'elles font, car elles possèdent seulement des systèmes de *sous-routines minimalistes* adaptés. Leurs évaluations représentationnelles internes sont déterminées en partie par les gènes, en partie par l'apprentissage, mais non par une activité libérée des contraintes environnementales immédiates. Par exemple, si tout nouveau problème arrive, leurs informations sont trop spécialisées et cloisonnées pour le résoudre. Un castor dans le désert ne survivrait pas.

À ce stade, notre bête n'a pas de «soi»; pour que la représentation de type «subjectif» («sel-morphic») ou conscient (humainement) émerge, il faut, selon Prinz, *deux étapes développementales* qui se succèdent dans le temps : primo, le développement de la capacité de re-présenter[87] ce qui n'est pas présent et de maintenir une telle re-présentation distincte de la perception directe (Prinz nomme cette capacité *double représentation,* elle serait la condition naturelle et innée du «soi»); deuzio, le développement de la posture interprétative qui fait des re-présentations le résultat d'une action individuelle; c'est-à-dire provenant de personnes. Prinz nomme cette étape l'*attribution de la source aux personnes,* qui serait la partie culturelle dans l'histoire du «soi».

2.2.2.1-Double représentation (histoire naturelle)

Premièrement, pour que les évaluations mentales cessent d'être enchainées au présent, qu'elles n'impliquent plus seulement quelques options prédéterminées de comportement face à un stimulus présent, un animal doit pouvoir rendre du contenu mental disponible de manière récursive pour d'autres types de traitement, actuels ou

[87] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

futurs. L'animal doit posséder la capacité de re-présenter, à savoir, de former des re-présentations de circonstances qui ne peuvent pas être perçues au moment présent.

Selon Prinz, ce type de contenu est d'abord apparu au sein d'associations sociales où se sont développées des formes simples de communication symbolique. Symbolique dans la mesure où des messages oraux réfèrent à des circonstances et plus particulièrement des circonstances au-delà de l'horizon perceptuel immédiat. Bref, Prinz semble assumer que les re-présentation nous ont d'abord permis de raconter aux autres ce que nous faisons ou ce que nous ferons. Maintenant, pour comprendre une telle communication symbolique, une architecture cognitive à deux niveaux est nécessaire. Une architecture qui peut traiter des re-présentations sur un plan et continuer à traiter l'information perceptuelle en même temps. Marcher et parler. Cette capacité d'exécuter des contenus perceptifs et des contenus re-présentés de manière juxtaposée permet un bond cognitif. Principalement elle permet la naissance du «soi» comme conséquence de l'attribution de la source des re-présentations à une personne.

2.2.2.2- *Attribution* des pensées à une personne

Prinz soutient que les re-présentations sont d'abord apparues à travers des communications, c'est-à-dire qu'elles sont d'abord déclenchées par la réception de messages verbaux, donc, induites de l'extérieur. Pour ces messages perçus dans l'environnement immédiat, il y a toujours quelqu'un et cette personne est la source perceptible du message.

Mais à partir du moment où certains peuvent raconter quelques chose à d'autres, ceux-ci peuvent très bien se raconter quelque chose à eux-mêmes. Ainsi, le système de double représentation peut produire des représentations depuis l'intérieur, telles que les souvenirs, les pensées, les fantaisies. Prinz nomme «pensées» toutes formes

de représentations mentales induites depuis l'espace interne. Maintenant, ces pensées sont des actes de représentation générés intérieurement qui ne sont pas accompagnés d'une perception immédiate d'un acte de communication. En d'autres termes, elles ne peuvent être attribuées à aucune source extérieure humaine dans la situation immédiate. Partant, d'où viennent les pensées?

La suggestion, à caractère spéculatif, de Prinz est de transférer le schéma pour interpréter les messages induits de l'extérieur aux pensées induites de l'intérieur aussi. Par conséquent, on fait remonter les pensées à une source présente dans la situation. Ici, rappelez-vous de votre tante lointaine qui discutait avec Jésus: certains font remonter les pensées à des voix: dieu, ange, ancêtre, esprit, daïmon, bref à une autorité personnelle qui est considérée comme ayant une présence invisible dans la situation immédiate. D'autres, par contre, situent la source des pensées dans une autorité personnelle autonome liée au corps de l'acteur : le «soi».

Ces deux solutions au problème de l'attribution de la source des contenus n'apparaissent pas simultanément dans l'histoire. La première est plus ancienne que la deuxième. De plus, différents facteurs entrainent le passage de l'un à l'autre: historique, sociologique, psychologique et politique. Selon Julian Jaynes cité par W.Prinz[88], ce passage de l'une à l'autre solution serait parvenu précisément durant l'antiquité. La trace historique de cet avènement est le personnage d'Ulysse à qui Homère a fourni une conscience interne. En effet, dans l'Odyssée, contrairement aux personnages de l'Illiade dont les pensées obéissent aux dictats des dieux (comme tous les autres avant), Ulysse a un «soi» et c'est ce «soi» qui pense et agit.

Ce développement du soi dans l'antiquité semble reposer sur des conditions politiques particulières. En effet, les sociétés qui attribuent leurs pensées à des voix

---

[88] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

produisent des hiérarchies dont les puissants légitiment leur pouvoir en proclamant qu'ils représentent sur Terre les divinités qui en sont la source ou qu'ils en sont les interprètes légitimes. C'est seulement lorsque le «soi» prend la place des dieux que de telles hiérarchies deviennent inutiles. Les organisations politiques autoritaires disparaissent au profit de nouvelles organisations qui fondent leur légitimité sur la volonté de la majorité des citoyens, en l'occurrence sur la volonté de sujets qui sont perçus comme étant autonomes. Dans le même sens, quelques dizaines d'années après Homère (à la fin du 6^{e} siècle av. J-C.), la démocratie apparait à Athènes.

Maintenant, dans ces sociétés plus individualistes, on n'attribue pas seulement la source des pensées à des «soi», mais aussi la source des actions. Précisément, le système de double représentation induit aussi une représentation indépendante de buts, à savoir, d'événements futurs potentiels qui sont désirés. L'action peut être planifiée de manière à réaliser des situations désirées.

Ici encore, il est possible d'attribuer la source des buts générés depuis l'intérieur à la volonté d'une personne d'autorité invisible : Dieu, daïmon, etc. Par contre, la solution des sociétés plus individualistes est de situer la source des buts dans des «soi» autonomes et libres. Partant, «les sociétés qui placent le «soi» à la position qu'occupe normalement les dieux, les rois, font naitre des agents autonomes»[89]. Cette attribution à une autorité personnelle est une construction générée par une certaine organisation sociale et politique et non un organe naturel de la pensée. C'est un concept que l'on projette sur un système de *double représentation.*

L'invention du «soi» autonome et libre n'est pas le fruit d'un grand philosophe. Les «soi» sont construits et maintenus par des échanges sociaux concrets, c'est-à-dire que l'on enculture les rejetons de ces sociétés dans une culture – une banque

[89] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003p.8

d'informations et de discours – qui conçoit l'être humain comme possédant une subjectivité, une conscience. Cette culture fournit une interface personnelle qui organise leur structure mentale[90].

D'abord, en Occident, la psychologie du sens commun, la posture intentionnelle dirait Dennett, permet de penser l'humain avec un «soi» en son centre qui demeure identique dans le temps et qui est le centre représentationnel de la pensée et le centre de prise de décision pour l'action[91].

Partant, l'éthique et particulièrement le système judiciaire des sociétés plus individualiste font reposer la responsabilité de l'action sur les individus précisément parce qu'ils les conçoivent comme autonomes, libres, comme possédant un «soi» qui est la source des pensées et des actions.

Dans ce sens, la psychologie du sens commun, qui est au fondement de la culture, a le rôle d'une institution sociale qui régule nos pensées et nos actions[92]. La validité de cette psychologie populaire a une importance capitale pour les sociétés individualistes. Leurs institutions juridiques, morales, politiques, familiales, etc. reposent sur elle. Partant, il ne faut pas s'étonner si dans notre culture elle a longtemps été intellectualisée par les savants, les philosophes et maintenant par les scientifiques[93].

---

[90] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

[91] Ibid

[92] Kosch and Leary, *A centery of psychology as science*, 2006

[93] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

3- Histoire de l'individualisme et histoire du concept

Nous avons soumis l'hypothèse à la fin de la section 1.1 de ce chapitre que puisque les concepts sont les ensembles de base de tout ce qui se trame dans l'espace interne personnel (le «soi») alors ils sont eux-mêmes des outils théoriques importants pour toute l'idéologie individualiste; nous avons proposé l'idée qu'un membre d'une société individualiste, lorsqu'il tente d'expliquer la nature des comportements humains, doit référer à cet espace interne implicite, décrire les phénomènes mentaux et leurs contenus, quelque chose comme les concepts, nécessaires aux pensées et aux actions individuelles. Or, le concept de concept émerge à la même époque, au même endroit, dans les mêmes conditions socio-culturelles que le «soi». Ces faits historiques corroborent notre hypothèse, à savoir que les concepts sont définis en fonction de leur participation aux structures sociaux-psychologiques individualistes.

D'abord, avant que le «soi» n'apparaisse, les sociétés traditionnelles sont organisées suivant une hiérarchie qui au sommet place un individu sur lequel repose le pouvoir. En fait, ce souverain forme l'intermédiaire entre les humains et le divin et parfois même ce souverain est une divinité incarnée ou un descendant d'une divinité. Par conséquent, le pouvoir de domination et plus particulièrement les décisions politiques émergent ou sont transmises par cet être qui en impose[94]. Dans le même ordre d'idée, lorsque que ces sociétés traditionnelles cherchent à comprendre l'organisation du monde en général, elles transfèrent leur solution perçue dans la sphère sociale et projettent des dieux, rois du ciel, qui sont la source de l'organisation du monde, comme les rois sont la source de l'organisation du groupe[95].

---

[94] W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

[95] J-P Vernant, *Mythe et pensée chez les Grecs. Études de psychologie historique*, Paris, François Maspero, 1965 ; rééd. Paris, La Découverte, 2007

Ensuite, lorsque, selon Jaynes[96], des cités sont en voie de démocratisation au 7e siècle av. J-C (cités oligarchiques dont le pouvoir est de plus en plus partagé) le «soi» apparait puisque ces cités fondent leur légitimité sur la volonté d'un grand nombre de citoyens. C'est-à-dire que le pouvoir repose de plus en plus sur des individus considérés comme possédant une certaine autonomie.

De même, au 6e siècle, le «soi» s'institutionnalise explicitement dans les nouvelles démocraties. En l'occurrence, le pouvoir de domination, le *kratos*, est déposé *au centre de la cité à égale distance de chaque citoyen*[97]. Ici chacun n'est dominé par personne. La source des décisions qui concernent le groupe n'est plus attribuée aux dieux, aux rois intermédiaires ou à toutes autres sources extérieures aux individus[98]. Le pouvoir de décision politique devient neutre et est exercé par la communauté elle-même à travers la discussion, l'argumentation, le dialogue, le débat. C'est la cité, ou plutôt les citoyens qui sont responsables des politiques. Ils en sont la source en quelque sorte.

Maintenant, si ce qui est *bien* politiquement ne correspond plus à des dictats extérieurs à la volonté des individus, mais est fondé sur les pensées de ces derniers, alors à quoi ces pensées correspondent-elles? Les grecques de l'antiquité apportent au moins deux réponses différentes à ce problème.

Primo, si les décisions politiques reposent sur le vote des citoyens à l'ecclésia, à savoir, si elles sont produites par les individus, alors ce qui est *bien* politiquement peut être compris comme relatif à la majorité. En d'autres termes, puisque les

---

[96] Jaynes, in W.Prinz, *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

[97] Vernant, *Mythe et pensée chez les Grecs. Études de psychologie historique*, Paris, François Maspero, 1965 ; rééd. Paris, La Découverte, 2007

[98] J-P Vernant, Ibid

pensées (ex.: *bien* politique) sont produites par l'espace intérieur des individus, elles sont relatives à ces derniers.

Ce relativisme politique est généralisé à toute la connaissance par les sophistes relativistes[99]. Pour Protagoras par exemple, les connaissances sont toujours relatives à l'individu dans la mesure où celles-ci sont induites empiriquement, à savoir, qu'elles arrivent dans l'espace interne personnel depuis le monde en passant par les sens de l'individu. Or, chacun a des dispositions sensorielles différentes, une expérience vécue différente, une perspective spatiale différente, une santé différente, etc. Partant, deux individus peuvent avoir chacun une connaissance basée sur des impressions qui peuvent être vraies pour chacun, mais contradictoires. La température peut être chaude pour un et froide pour l'autre. Pourtant, ce qui est contradictoire ne peut pas être vrai. Par conséquent, si la vérité, pour les relativistes, n'entre pas dans l'individu par impression sensible, il est vain de la rechercher. Le monde en est un d'apparences.

Mais certains, les philosophes comme Socrate, veulent dépasser cette relativité individuelle et saisir les essences des choses au-delà des impressions et des opinions[100]. Ils veulent saisir les définitions universelles, éternelles et non-contradictoires. Car, comme il est difficile d'imaginer que 2+2=4 soit une connaissance relative aux individus, les philosophes chercheront la définition des idées telle que celle du *bien universel* au-delà des exemples particuliers perçus et pensés par un individu. Mais, si les essences ou les définitions universelles telles que celles en mathématiques ne sont pas perceptibles dans le monde naturel, sensible, alors comment les saisir et où existent-t-elles? La réponse de Socrate, et de Platon qui écrit les dialogues socratiques, est que ces essences sont situées en dehors

[99] Platon, Théétète, 4e av. J.-C.
[100] Platon, *La République* , Livre 7.

du monde sensible, dans un monde intelligible et surnaturel[101]; partant, elles sont connaissables grâce à l'âme (qui seule possède d'emblée cette nature surnaturelle et intelligible (voir la réminiscence)).

En lançant une quête des essences, des définitions universelles, comme ultime source ou source idéale des pensées, Socrate lance sans le savoir explicitement l'étude des concepts. «Avec Socrate nous nous trouvons au seuil d'une révolution de la pensée : la découverte du concept»[102]. Toutefois, si Socrate est le premier à rechercher des définitions universelles (du bien et des vertus en général), il ne s'est pas interrogé sur la nature de l'universel comme tel.

Certains sont peut-être tentés ici de critiquer l'idée que Socrate et Platon sont les instigateurs implicites de l'étude des concepts (objet habituellement internaliste), dans la mesure où ils semblent plutôt réfractaires à la révolution individualiste qui se trame en Grèce puisqu'ils expulsent la source véritable des pensées en dehors de l'individu, dans un monde surnaturel. Il est vrai que le concept de concept dans l'histoire qui suit est plus souvent qu'autrement une entité interne. En fait, si Platon expulse la source des pensées hors de l'individu, c'est *grâce à* (en réaction à) l'individualisme de la démocratie[103]. Le point à retenir est qu'il ne le fait plus de la manière traditionnelle. Il n'extériorise pas la source en la projetant dans un dieu, mais dans des Idées intelligibles et rationnelles.

Or, ces Idées platoniciennes sont annonciatrices des concepts. Comme les concepts qui apparaissent après Platon, les Idées sont des entités à caractère universel, nécessaire et abstrait. Maintenant, contrairement aux concepts, elles sont elles-mêmes des objets de connaissance, saisissables immédiatement, alors que les

---

[101] Ibid.

[102] J-P Vernant, *Mythe et pensée chez les Grecs. Études de psychologie historique*, Paris, François Maspero, 1965 ; rééd. Paris, La Découverte, 2007

[103] Popper, *La société ouverte et ses ennemis*.

concepts forment un type de re-présentation, c'est–à-dire qu'ils sont des objets qui se rapportent à quelque chose d'autre qu'eux-mêmes. En effet, l'on s'entend traditionnellement pour décrire les concepts comme des re-présentations qui permettent à l'esprit de regrouper plusieurs membres d'une même catégorie, dans la mesure où ils sont des ensembles circonscrits par des critères nécessaires et suffisants; critères qui, aujourd'hui, ne sont pas compris comme des propriétés des choses représentées.

Les concepts se distinguent aussi de l'Idée en ce qu'ils sont habituellement le résultat d'un acte de conceptualisation; ils proviennent d'un travail de généralisation opéré par l'esprit. Par conséquent, les concepts existent dans des espaces intérieurs personnels. À l'inverse, c'est lorsqu' on s'engage sur des objets externes, qui ont le même genre de nature, qu'on discute comme Platon d'Idée, ou plus tard d'Universaux, de catégories, etc. Bref, les concepts ont de particulier qu'ils sont un mode de connaissance, un outil intellectuel interne par lequel un sujet peut penser des choses différentes dans le monde, ce que ne sont pas encore les Idées.

C'est Aristote qui conçoit les concepts comme possibilité empirique des essences. Selon ce dernier, ami de Platon mais encore plus de la vérité[104], rien n'est dans l'intelligence qui ne provienne des sens[105]. Il partage ainsi la thèse empiriste des relativistes et refuse celle d'un monde surnaturel et celle de la réminiscence. Mais, si la connaissance provient de l'expérience sensible, et que chacun fait des expériences différentes (relativisme), comment saisir la vérité, les définitions universelles? Aristote immanentise les essences au monde: les choses perceptibles sont composées d'une matière et d'une forme. Les sens font entrer les attributs physiques d'une chose dans l'espace interne et l'intellect abstrait la forme

---

[104] Aristote, Éthique à Nicomaque, CEC inc., 2010, I, 4.
[105] Aristote, *La Métaphysique* (trad. Pierron et Zévort), Livre Premier. http://fr.wikisource.org.

universelle à partir de ces attributs physiques. Le résultat de ce travail interne ce sont les concepts.

Avec les concepts, Aristote se positionne entre Platon d'un côté et Protagoras et Gorgias de l'autre. En effet, pour les relativistes, connaître se limite aux impressions sensibles, aux apparences; il n'y a pas de vérité universelle, alors que pour Platon, connaître c'est s'éloigner du sensible pour saisir les Idées pures. Or, Aristote réussit à expliquer la conceptualisation de l'intelligible au sein du sensible.

Suite à Aristote, il semble possible d'observer une tendance vers une internalisation ou une humanisation du caractère universel ou général des concepts. En effet, Aristote situe les formes universelles (abstraites lors de la conceptualisation) dans les substances externes. Elles sont immanentes au monde naturel. Or, plus les organisations socio-politico-culturelles favorisent l'individualisme dans l'histoire, plus il semble que la source des concepts sera relative à l'individu[106].

Par exemple[107], aux 13^{e} et 14^{e} siècles, l'individu devient une catégorie fondamentale du droit un peu partout en Europe et particulièrement en Angleterre où Ockham opère en parallèle une révolution épistémologique avec son nouveau nominalisme. En d'autres termes, la révolution sociale qui amorce une libération de l'individu face aux obligations communautaires traditionnelles apparait en même temps que l'idée que seules les individualités existent: il n'y a aucune entité de nature universelle en dehors de l'espace interne tels que les Universaux ou les autres substances aristotéliciennes. Pour Ockham, l'esprit n'extrait aucune forme générale qui serait

---

[106] Alain Laurent, Histoire de l'individualisme, Presse universitaire de France, 1993.

[107] Le but n'est pas de couvrir l'histoire de manière exhaustive, mais de faire ressortir la relation entre l'individualisme et l'intériorisation des *idées*. Nous croyons qu'il est possible de saisir abstraitement des phénomènes historiques, un peu à la manière de Google Ngram, en synthétisant 2000 d'histoire dans un cours portrait. Comme l'idée du big bang synthétise le phénomène d'erxpension de l'univers. Ainsi, par exemple, nous ne discutons pas des hellénistes, d'Augustin, de Descartes, de Lock, etc., qui ont tous une énorme influence sur cette histoire.

immanente aux choses. Les concepts sont ici des entités singulières de nature physico-psychique qui servent de signes à une pluralité de choses singulières et extérieures qui se ressemblent. Bref, les concepts sont identiques à des états mentaux internes qui remplissent la fonction de signes généraux des choses externes.

Pour poursuivre notre petite image de l'internalisation historique des concepts, on peut se rappeler que durant la Renaissance, l'homme protestant devient un sujet spirituellement autonome, voire autosuffisant. Ceci se traduit, entre autres, par des droits et des libertés inaliénables comme ceux du *Bill of Rights (1690),* par un libre marché, par l'idée d'une nouvelle démocratie, etc. On voit apparaitre l'état de droit moderne qui a pour but d'établir la société idéale à partir de l'isolement de l'individu[108]. En bout de piste, c'est la *Déclaration des droits de l'homme et du citoyen* qui sacralise le succès de cet individualisme : *«« le but de toute association politique est la conservation des droits naturels et imprescriptibles de l'homme » qui revient à dire que la « liberté» et la « propriété» sont des droits appartenant à la nature de chaque individu et qu'ils ne proviennent pas de l'appartenance à la société»*[109].

Le concept de concept chez Kant reflète bien l'intellectualisation par les savants de cet individualisme amplifié. Pour Kant, on n'abstrait pas (pas plus qu'Ockham) une forme universelle du monde à partir des choses senties; on ne peut abstraire qu'une représentation générale (un concept empirique ou a posteriori) de ce qui a de commun entre plusieurs individualités. Mais à l'instar du problème général de l'induction (Hume), ce type de concept est toujours susceptible d'être contredit par l'expérience. Par conséquent, il ne répond pas aux critères classiques du concept de concept. En fait, le «centre» de la connaissance universelle et nécessaire est le sujet

---

[108] Alain Laurent, *Histoire de l'individualisme*, Presse universitaire de France, 1993
[109] Ibid.

connaissant, et non une réalité extérieure. Ainsi les conditions de possibilités d'une telle connaissance sont à chercher à l'intérieur même de l'espace interne individuel.

Pour Kant, la réalité extérieure est inconnaissable en soi, elle est toujours pensée en fonction de contraintes logico-cognitives, à travers la structure de l'entendement. Or l'entendement possède des types de jugements et des concepts a priori qui ont pour fonction *d'ordonner des représentations diverses sous une représentation commune*[110]. Ce travail en est un de synthèse par lequel l'entendement lie la diversité sensible a posteriori en des concepts a priori.

Soulignons que l'entendement ne classe pas, il ne cherche pas parmi des représentations particulières à reconnaitre des caractéristiques objectives et communes à une classe, il impose plutôt ses propres critères, ses lois, aux objets. C'est-à-dire que les objets qui entrent dans l'expérience sont différents des représentations singulières qui en sont affectées et ces dernières sont différentes encore des concepts qui les ordonnent; or les pouvoirs de représentation (sensible) et de conceptualisation ont des règles et celles-ci conditionnent la connaissance des objets. Les dispositions individuelles (les concepts) *imposent une manière de voir*[111].

En quelque sorte, avec les concepts, Kant intériorise la forme universelle immanente au monde d'Aristote. Les dispositions de l'espace interne personnel (le «soi», qui est en fait un construit social) devient l'ultime source du caractère universel et nécessaire de la connaissance. Cette façon de poser le problème est nouvelle dans la mesure où elle lie la question de la connaissance à celle du savoir philosophique et scientifique de façon fondamentale. Concevoir que l'esprit impose des formes et une structure à ce qu'il connait (et que, de l'autre côté, les objets imposent des

---

[110] Extraits de textes de Kant, Critique de la raison pure. Trad Alain Renaut. GF Flamarion 2001 Cours Prof. Jean- Guy Meunier

[111] Ibid.

limites à ce que l'on peut connaitre) marque un tournant qui a encore des échos en psychologie et en science cognitives.

Sciences cognitives[112]

« Toujours depuis Kant, ..., a existé parmi les philosophes une tendance que je considère comme illusoire, la tendance à admettre que la description du monde est influencée outre mesure par des considérations dérivées de la nature de la connaissance humaine »[113]

Suite à Kant, la tendance individualiste en est une de masse et la notion d'individu *se concrétise*[114]. L'industrialisation et la division du travail, l'économie libérale et la notion d'espace privé, le droit de contrat entre personne privée, l'extension du suffrage universel, etc. font en sorte que petit à petit l'individu agit au quotidien selon son bon vouloir. En fait, une foule d'hommes au 19e siècle, surtout urbains, désirent davantage vivre leur vie selon leur propre fantaisie que toute autre génération précédente. Ils désirent être différents de leur famille, affirmer leurs idées, aimer en dehors du mariage de raison, créer en dehors des canons culturels, etc. Par exemple, tout le courant romantique exprime bien cette idée.

Cette liberté grandissante complexifie l'intériorité des individus dont les processus psychologiques demandent à être décrits et expliqués. Or, si au 18e siècle on s'entend encore sur l'impossibilité d'expliquer scientifiquement ou empiriquement l'esprit comme on le fait pour le monde extérieur, le 19e siècle marque le début de l'étude de cet espace interne avec des critères d'objectivité, des méthodes et une rigueur propres aux sciences (Lotze, Brentano, Helmholtz ou Wundt).

---

[112] Nous sommes conscients encore une fois que nous passons outre des moments historiques essentiels à une véritable histoire de l'internalisation des concepts. En l'occurrence nous avons décidé de ne pas discuter de l'anti-psychologisme au 19e siècle dans la mesure où ce n'est pas l'objet principal de ce mémoire et partant pour un souci d'économie d'espace. Toutefois nous y reviendrons brièvement dans le chapitre 4.

[113] Russell, *Human Knowkledge,* (1948)

[114] Alain Laurent, L'histoire de l'individu,

Rapidement, plusieurs innovations scientifiques au 19e et au début 20e en neurosciences et en logique formelle semblent permettre à la psychologie et d'autres sciences de la cognition d'assurer la pérennité du paradigme internaliste en posant un discours objectif sur l'espace psychologique interne et ses représentations internes.

C'est que pour être une véritable science, la psychologie, en plus de son explication logico-conceptuelle, doit pouvoir dépasser les postulats subjectivistes en expliquant de manière naturaliste les processus élémentaires de l'espace psychologique interne, du «soi» et les entités de base qu'il utilise (les concepts). Or, des innovations majeures provenant des sciences informatiques et des neurosciences apportent des solutions dans ce sens.

D'abord, les neurosciences tentent d'expliquer matériellement certains processus psychologiques, entre autre par la description du système nerveux et de ses constituants, les neurones. Parallèlement, les sciences informatiques émergeantes trouvent le moyen de faire le pont entre les processus logiques et ceux causaux (matériel) en développant les machines computationnelles.

La convergence entre la psychologie, les neurosciences et l'informatique ouvre un nouvel horizon théorique : la possibilité d'une explication à la fois matérielle et logique (conceptuelle) des entités mentales. Cette rencontre donne naissance à ce que l'on peut appeler aujourd'hui les sciences cognitives qui aux dires de certains est une véritable *révolution cognitive*[115]. Un programme de naturalisation du «soi» comme espace interne individuel.

Hic et nunc

---

[115] Gardner, H. (1987). *The mind's new science: a history of the cognitive revolution.* New York: Basic Books

C'est ici que nous finalisons le court portrait historique qui nous permet maintenant de mieux comprendre et expliquer les causes qui poussent certains, même parmi les plus savants, à limiter leurs recherches de l'origine des concepts et leurs contenus à des intérieurs crâniens. En effet, ce recul historique montre comment le paradigme internaliste en sciences cognitives est l'héritier contemporain d'un internalisme beaucoup plus ancien qui lui est engendré par des impératifs historico-culturels. Il est même maintenant plus facile de comprendre, avec cette perspective bimillénaire, les raisons qui ont mené à l'essor en général des *sciences cognitives cartésiennes* dans nos démocraties, ce qui est bien illustré par la déclaration de « la décennie du cerveau» de G. Bush en 1990[116] et qui a mené à des investissements massifs dans le domaine. Mais, si nous comprenons mieux les motifs de l'internalisme, nous comprenons mieux du même coup que le programme de *cérébralisation* de la source des concepts est en partie un phénomène contextuel et relatif à un type d'organisation sociale. En effet, Prinz soutient de manière convaincante que le «soi» mental est apparu pour solutionner un problème d'attribution de la source des représentations et que certains mécanismes d'interactions sociales et certains discours maintiennent et préservent l'institution du «soi» mental.

Par conséquent, si les sciences cognitives sont une entreprise de naturalisation de l'espace psychologique interne, mais que plusieurs caractéristiques de cet espace telles que la liberté et l'autonomie sont en grande partie un construit théorique nécessaire au fonctionnement de nos sociétés individualistes (dans la mesure où il permet d'attribuer la source des représentations et des pensées en général aux individus), alors toute l'entreprise est fondée sur une idée reçue qui est vouée au quiproquo. Ce n'est pas parce qu'un certain type d'organisation sociale invite à projeter des choses qui lui sont utiles sur des individus, des «soi» absolument responsables de leurs pensées, que ces choses existent tels quels dans ces mêmes individus.

---

[116] http://www.mels.gouv.qc.ca/sections/viepedagogique/numeros/117/vp117_49-52.pdf

Qu'en est-il alors des concepts? Si les «soi» sont en partie des fictions utiles aux sociétés, comment définir leurs contenus? Entendons-nous, les individus acquièrent, reproduisent et améliorent à un certain point de véritables représentations, même généralisantes. Le système de double représentation de Prinz, par exemple, correspond peut-être aux mécanismes réels d'un organe inné et essentiellement biologique. L'argumentation de Prinz nous permet seulement de croire qu'il n'y a qu'une partie des concepts qui est fictionnelle. En l'occurrence l'attribution de leur source à des «soi» libres et automnes. Mais alors, si c'est le cas, la thèse de l'interdépendance théorique entre les processus mentaux internes et les concepts(p.63) perd de sa pertinence. Les attributs fonctionnels des concepts dans les théories plus empiriques qui relèvent directement de la nature des processus mentaux perdent leur valeur ontologique (dans la mesure où nous avons montré que l'explication de l'espace interne à l'aide des processus ou des mécanismes cognitifs était la version *dernier cri* de l'internalisme bimillénaire qui a produit le «soi»). Et si la nature des concepts n'est plus uniquement relative à la nature des processus mentaux, alors la nature des concepts n'est plus exclusivement relative à celle des processus cérébraux propres à la cognition supérieure (puisque les processus mentaux sont identiques aux processus cérébraux).

Or, quelle est la nature des concepts?

4-Où vaporiser[117] les concepts

Quelle sorte de chose tente-t-on d'expliquer avec le concept de concept une fois que l'on a dit qu'il ne se réduit peut-être pas aux assemblées de neurones à l'intérieur

[117] Nous faisons allusion à l'expression de D. Dennett dans ses conférences sur *Breaking the spell :* Si tout est naturel et matériel, où doit-on appliquer le vaposisateur pour arroser les concepts.

d'un crâne? Si l'on veut vaporiser ce genre de substance, où et sur quoi faut-il le faire? Quelle est dans ce cas la nature véritable de la relation qu'entretiennent les concepts et les populations de cellules humaines si cette relation est dans une certaine mesure une fiction créée et maintenue par nos sociétés individualistes? Quelle part des concepts retrouve-t-on réellement tels que définis par les théories, en observant un organisme humain?

Dans un sens, ces questions sont également épistémologiques. Prises dans leur généralité, elles consistent à s'interroger sur la manière dont le concept de concept représente les concepts, sur la manière dont ce concept représente le monde. De la même manière, lorsqu'on se demande si le concept «cheval vapeur» ou «Joe six pack» représente adéquatement une réalité, on se demande du même coup s'il y a une part d'interprétation (fiction) qui demeure relative à ces représentations. Ici, par exemple, on peut mesurer la puissance d'un moteur en «cheval-vapeur» ou en «watt» sans présupposer que les qualités mesurées appartiennent, absolument, à l'ontologie de l'objet (moteur). Peut-on concevoir le concept de concept de la même façon?

Nous avons proposé l'hypothèse que les concepts n'existent pas non plus, objectivement, tels que décrits. Comme pour les chevaux vapeurs et les watts, il est possible et cohérent de décrire les représentations généralisantes stockées dans une mémoire en termes d'exemple, de prototype, de théorie, d'idéal, de définition, etc., sans présupposer que toutes les qualités décrites appartiennent à l'ontologie de cet objet. En réalité, c'est en partie ce que fait implicitement chaque chercheur de la communauté en poursuivant un débat entre les principales théories. C'est-à-dire que chacune des théories est supposée être la seule à décrire ce qu'est un concept, alors que dans les faits plusieurs théories distinctes, en compétition, sont en même temps compatibles avec une grande part des comportements observés. Il ne semble pas y avoir (encore ?) de fait observable qui permettrait d'éliminer ou de déclarer comme

vainqueur une théorie plus qu'une autre de manière définitive. Ainsi, à la manière de Ryle, il est possible que le vocabulaire des concepts soit simplement une manière de décrire une certaine définition de la cognition supérieure réelle. Néanmoins, même si la signification du concept de concept ne se réduit pas absolument à son extension, cela n'implique pas nécessairement qu'il faille l'éliminer. L'idée des niveaux de descriptions telle que D. Dennett la développe le montre bien. Il est possible dans certains cas de discuter d'un même phénomène cognitif à partir de différentes postures. Par analogie, on peut décrire le comportement de notre ordinateur en se référant aux représentations sur l'interface: «mon ordi veille» ou « Thémistocle est jeté dans la corbeille», on pourrait aussi décrire cette action en termes fonctionnels, ou la décrire en termes matériels. Si l'on utilise un exemple plus proche de la véritable cognition, on peut discuter de la mémoire humaine à un niveau représentationnel comme étant notre capacité de se rappeler des expériences passées; on peut en discuter depuis un niveau cérébral comme étant un mécanisme qui implique le cortex, plus précisément le lobe préfrontal et l'hippocampe; on peut en discuter à un niveau cellulaire comme étant l'encodage de représentations sur des assemblées de neurones et des patterns de connexions synaptiques; ensuite à des niveaux moléculaire et atomique. Soit, D. Dennett soutient qu'il y a des *parties* ou caractéristiques de phénomènes qui sont plus facilement, voire uniquement, connaissables en adoptant la bonne posture.

Par exemple, il serait difficile de comprendre pourquoi une foule hurle lorsqu'une rondelle s'introduit dans un filet en n'observant la scène qu'au niveau moléculaire. On risque de mieux comprendre et plus rapidement le comportement des individus en faisant l'hypothèse qu'ils possèdent les concepts «but», «victoire», etc. La même chose pour un meurtrier, il est plus facile d'expliquer son geste en lui attribuant des pensées, des intentions, etc. Par contre, pour comprendre et prédire la trajectoire de la rondelle ou de la balle de fusil, une description physique en termes de masse, d'énergie, de vélocité, etc. est avantageuse.

Ainsi, les concepts décrits par les sciences de la cognition possèdent une certaine réalité objective dans la mesure où ils peuvent décrire de manière abstraite, générale et synthétisée certains phénomènes cognitifs. En l'occurrence, le concept de concept permet d'expliquer à un certain niveau d'abstraction comment un individu (somme de cellules) réalise certaines tâches expérimentales reliées à la connaissance: catégorisation, induction, déduction, etc. Mais dans un autre sens, puisqu'on ne peut discerner certaines réalités qu'en adoptant la bonne posture, les concepts n'existent que relativement à une théorie et pour celui qui s'engage sur cette dernière. En l'occurrence, certains phénomènes cognitifs sont connus seulement si la théorie des concepts utilisée pour interpréter quelques tâches expérimentales est celle des prototypes, d'autres phénomènes cognitifs sont bien expliqués si les concepts sont compris comme des exemples et d'autres phénomènes encore sont saisis si les concepts sont compris comme des théories. Partant, l'existence des concepts est en partie relative à une hypothèse sur le fonctionnement cognitif réel de certains organismes vivants, de même leur existence est en partie relative (sa partie qui n'est pas adéquate à la réalité représentée est relative) aux individus qui soutiennent cette hypothèse. Dans ce sens les concepts sont des fictions utiles.

Bref, pour répondre explicitement: de quel genre de substance sont les concepts? Des informations inscrites dans la tête? Une structure électro-chimique? À l'instar de Dennett, il ne faut pas s'engager là-dessus directement, mais il faut plutôt concevoir le concept de concept comme une fiction de théoriciens dans une posture représentationnelle. Une fiction opérative et efficace à en croire l'histoire. Une fiction comme le concept «Canadien moyen» ou «chevaux vapeurs». Bref, il faut traiter les concepts comme des abstractions utiles pour décrire et expliquer les processus cognitifs supérieurs de populations spéciales de cellules elles-mêmes organisées en sociétés individualistes, c'est-à-dire en sociétés qui attribuent la source des pensées, et donc des concepts, aux individus-cerveaux.

En d'autres termes, nous ajoutons une couche de relativisme sur les théories des concepts. Nous arrivons en bout piste avec une description où les concepts sont des fictions utiles; d'abord intellectualisées par un paradigme internaliste lui-même en partie fictif: dans le fait qu'il attribue la source des concepts strictement à des espaces psychoneurologiques internes, des «soi» *cérébralisés*; qui plus est, cet internalisme est produit et maintenu à son tour par les impératifs fonctionnels de sociétés individualistes.

Cette description montre bien pourquoi les explications du courant dominant internaliste dans les *sciences cognitives cartésiennes* aboutissent dans un cul-de-sac, à savoir, qu'ils ne fournissent pas d'explications sur l'origine des concepts à partir de zéro. Ce programme tente de matérialiser une fiction en cherchant la source des concepts dans l'espace psychoneurologique interne. Or, le «soi» n'est justement pas un organe et les concepts ne se réduisent pas à des mécanismes cognitifs individuels.

CHAPITRE IV

COGNITION DISTRIBUÉE

1- Théorie duelle: représentation interne et représentation externe

Rappelons que l'objectif du mémoire est de développer une définition du concept de concept qui soit cohérente avec les principes matérialistes et naturalistes en général. Corollairement, nous cherchions une définition qui explique la fabrication d'un concept à partir de zéro. Notre thèse est que les approches psychologiques ou, plus largement, internalistes ne peuvent y arriver précisément parce que leur objet d'étude, l'individu (pris individuellement) n'est pas suffisant pour construire des représentations aussi complexes. Un des fils conducteurs de l'argumentation jusqu'ici est le suivant.

Au premier chapitre, nous avons montré que les concepts sont un type de représentation qui a la particularité d'être généralisante. De plus, il semble aussi admis par un nombre non négligeable d'auteurs en sciences cognitives qu'il existe au moins deux (sous)types de représentations généralisantes. Par exemple, selon Piccinini et Scott, il y a deux types de concepts, ceux explicites et ceux implicites. Précisément, il y a les représentations qui nécessitent la compréhension et l'utilisation d'une langue naturelle et les autres. Les premières sont observables seulement chez les humains capables de parler et les autres sont partagées par la

plupart des animaux. Les premières sont utilisées dans des processus de compréhension de langage, d'inférence linguistique et de combinaison lexicale. Les dernières de type implicite sont utilisées dans des processus de discrimination, d'inférence non-linguistique et de catégorisation. Les *exemples*, *prototypes* et *théories* (theory-theory) se résumeraient toutes à ces représentations généralisantes mais non-linguistiques.

D'autres encore comme Poirier et Beaulac semblent en accord avec cette discrimination et s'engagent davantage en soutenant que les représentations généralisantes de type2 (+ou – explicites) prennent la forme de définitions. En effet, selon eux, on peut distinguer les concepts utilisés ou réalisés par des processus automatiques et ceux utilisés ou réalisés par des processus réfléchis. Le premier type de représentations (type 1), qui regrouperait les *espèces* de Machery, serait utilisé dans des processus automatiques, alors que le deuxième type de représentations (type 2), utilisé dans des processus réfléchis, regrouperait des concepts plus classiques nécessitant du langage: des définitions par exemple.

Nous partageons l'idée qu'il existe vraisemblablement, au plan théorique, deux types de représentations générales qui peuvent être distinguables par leur participation ou non aux opérations linguistiques et réfléchies. Toutefois, on ne devrait peut-être pas discuter de différents types de concepts, mais seulement de différents types de représentations généralisantes desquels le type2, uniquement, remplirait la fonction propre aux concepts.

Premièrement, les trois sous-espèces de représentations généralisantes de type1 prises isolément n'expliquent pas le phénomène de conceptualisation. En effet, aucune des trois théories formulées[118] ne décrit ni n'explique à elle seule tous les phénomènes cognitifs observés et associés à l'utilisation de concepts. C'est-à-dire

[118] Celles discutées par Machery

que ni la théorie des *exemples*, ni celle des *prototypes*, ni celle des *théories* ne parvient à prédire toutes les variantes de jugements catégoriels, ni toutes les variantes de raisonnements inductifs ou déductifs qui sont relatifs au concept de concept. Pour certains, c'est un argument pour soutenir qu'il y a plusieurs sous-espèces de concepts. Toutefois, si l'on tient compte tant de l'architecture qui les soutient que des processus qui leurs sont associés[119], ce n'est pas le cas. En effet, les trois représentations généralisantes suivent une chaine référentielle lors de leur acquisition, ils correspondent en fait à trois étapes d'un même apprentissage hiérarchisé (Weiskopf, McClelland). Par conséquent, il se pourrait fort bien qu'il n'existe qu'une seule espèce de structure correspondant aux représentations de type1 dans la mémoire à long terme (du cortex)[120], un complexe de représentations superposées, duquel différents niveaux d'informations pourraient être utilisés, selon le contexte, pour des traitements futurs. Dans ce sens, il nous semble qu'il n'existe pas différents concepts de type1, mais différents niveaux de représentations, de complexités différentes, qui participent à un complexe représentationnel généralisant de type1.

Deuxièmement, les représentations de type1, même lorsqu'elles forment un complexe, ne satisfont pas les critères d'une définition *riche* du concept de concept. Il existe traditionnellement une distinction entre les percepts qui jouent le rôle de représentations généralisantes plus simples, telles que des classes, et les concepts qui participent à des opérations supérieures, qui permettent de connaitre plus loin que ce qui est possible avec la perception. Pourquoi en faire fi? En effet, depuis les présocratiques, en passant par Aristote, les médiévaux, Descartes, Kant, il existe une conception de la connaissance qui divise le travail de l'esprit en plusieurs facultés[121]. De façon générale, on distingue une première faculté relative à la perception qui est

---

[119] Revoir la section sur McClelland, p.48

[120] James L. McClelland, *Why there are Complementary Learning Systeme in the Hippocampus and Neocortex*, 1994

[121] Hardy-Vallée, *Quand penser c'est faire, Les concepts, devenus naturels.* Uqam. 2003

souvent plus passive, voire automatique; et une deuxième propre à la raison qui, elle, est plutôt réfléchie et consciente. Les concepts appartiennent à la deuxième. Par exemple, Descartes montre que l'on peut décrire un morceau de cire par ses attributs perceptuels : couleur, goût, odeur, forme, mais que ces caractéristiques ne nous fournissent pas le concept de cire. Si on chauffe le morceau, ce dernier perd ses attributs perceptibles mais demeure de la cire. Descartes en conclu que pour posséder le concept, la raison ou l'entendement doit opérer un travail supplémentaire sur les percepts, il opère «une inspection de l'esprit»[122].

Pourquoi aujourd'hui, si les auteurs qui discutent de théories duelles évoquent déjà, à l'instar d'une longue tradition, les caractères réfléchis, explicites et linguistiques des concepts (type2), ils subsument tous types de représentations généralisantes sous le concept de concept. Les caractéristiques typiques de ce dernier ne sont pas anodines, mais participent à former des représentations toujours éminemment singulières. En effet, si les *prototypes* et les *théories* forment un complexe de représentations généralisantes qui émerge de processus perceptuels ou catégoriels implicites, c'est-à-dire mettant en relation plus ou moins directe un individu sans réflexion et sans langage avec son environnement, peut-on espérer la création d'un complexe représentationnel d'une richesse telle que celui de «boson», d'« algorithme», de «fractal», de «courriel», etc. Les mécanismes propres aux représentations de type1 ne suffisent pas. Ces représentations plus proches de la perception participent sûrement à la conceptualisation, elles forment la matière première, mais elles doivent faire l'objet d'une deuxième transformation, c'est-à-dire qu'elles doivent être réfléchies et converties sous une forme linguistique pour atteindre la *richesse* des concepts. (Nous verrons un modèle connexionniste qui met en interaction des mécanismes perceptuels individuels et langagiers).

---

[122] Dsecartes, R., in B. Hardy-Vallée, *Quand penser c'est faire, Les concepts, devenus naturels.* Uqam. 2003.

De plus, si les concepts sont plus *riches* et complexes que les représentations généralisantes de type1 au point de les subsumer sous un concept différent, en l'occurrence celui de concept, c'est qu'ils participent à ce qui constitue la spécificité humaine, à savoir la capacité de traiter des représentations internes par une activité libérée des contraintes environnementales immédiates. C'est-à-dire que les concepts doivent pouvoir rendre leur contenu disponible de manière récursive pour d'autres types de traitement, actuels, futurs, imaginaires, fictifs, etc. Par exemple, si on demande à des enfants de représenter des choses comme des maisons avec des ailes ou des hommes à deux têtes, il s'avère que seuls les enfants qui possèdent un certain niveau de langage sont capables de le faire[123]. Nous avons vu au chapitre 3 que ce type de travail conceptuel implique des représentations qui réfèrent à des circonstances au-delà de l'horizon perceptuel immédiat et que seuls les représentations qui peuvent participer à une communication symbolique en sont capables.

Dans ce sens, on peut soutenir que les représentations généralisantes qui participent sous une forme linguistique à la réflexion dans l'espace virtuel interne d'un individu humain possèdent des propriétés que ne possèdent pas les représentations de type1, des propriétés qui rendent les premières éminemment plus sophistiquées au plan cognitif que que les dernières.

Troisièmement, un autre argument qui nous pousse à croire que l'on devrait distinguer les représentations généralisantes de type1 des concepts repose sur le problème de l'origine du contenu et partant, de la supervision de l'apprentissage. Nous avons montré au chapitre 2 qu'un individu seul devant un environnement contrôlé, surtout celui d'enfants en développement, semble pouvoir (re)construire des représentations généralisantes simples sans trop de rétroactions (mis à part

[123] Karmiloff Smith (1992), cité dans Meunier J-G, *Trois types de représentations*. Publication du LANCI. UQAM. 2002. www.unites.uqam.ca/lanci/

l'environnement contrôlé). Ceci appuie l'idée qu'il existe des mécanismes internes ou individuels de catégorisation. Toutefois, nous avons montré aussi dans le même chapitre que ces mécanismes atteignent rapidement leurs limites. C'est-à-dire, que plus les contenus des représentations sont complexes, plus la rétroaction semble nécessaire. Qui plus est, il semble y avoir un fossé assez large entre les représentations de type1 et celles de type2 si on les étudie sous cette posture. Par exemple, un touriste qui débarque à Montréal un jour de cueillette du recyclage peut sûrement réussir à classer, seul, tous les bacs verts sous la même catégorie avec succès, et ce, sans véritablement savoir ce que c'est. Mais les concepts «recyclage», «contenant» ou «vert» dans la tête de ce même touriste possèdent une ontogenèse très différente. Personne aujourd'hui ne peut se targuer d'avoir créé ces concepts *depuis zéro.* De la même façon, un *enfant sauvage* ne pourrait pas créer ces concepts au moyen de ses seuls mécanismes de catégorisation relatifs aux représentations de type1.

Si l'on tient compte de leur origine, il est physiquement impossible d'expliquer à l'aide d'un seul individu comment le contenu des concepts (les représentations généralisantes de type2) peut être créé, alors qu'il est rationnel de penser qu'un individu puisse (re)construire des exemples ou même des prototypes avec beaucoup moins d'interaction sociale. Partant, nous croyons que si l'on possède comme postulat la croyance que tout le travail de conceptualisation, depuis l'origine d'un concept, peut ou doit être incarné sur un individu (réseau de neurones, cerveau, robot, etc), il est alors avantageux de réduire au maximum les concepts à des percepts. En effet, s'il est impossible d'expliquer la création des contenus de concepts à l'intérieur d'une théorie internaliste, mais qu'il est souhaitable de maintenir cette dernière, il faut assujettir le concept de concept soit en faisant fi du problème de l'origine des contenus ou, comme Churchland, Machery et plusieurs autres, en ignorant les concepts de type2. Toutefois l'évitement n'est pas une solution aux problèmes de l'origine des contenus et de la supervision de

l'apprentissage. La conceptualisation *bottom-up* que tentent d'expliquer les modèles internalistes sera toujours handicapée sans une explication du travail *top-down* opéré par des agents extérieurs.

Soit, il ne manque peut-être pas grand-chose pour résoudre le problème de l'origine des concepts et du coup celui de la supervision de l'apprentissage. Peut-être que le principal obstacle n'est que la posture interprétative internaliste adoptée par plusieurs auteurs plus ou moins orthodoxes. En effet, si les véritables concepts (+ ou – représentations de type 2) ont des fonctions linguistiques, il semble possible de choisir un point de vue plus externaliste ou sociologique. Le langage peut peut-être fournir une explication naturelle et matérielle du travail conceptuel effectué à un autre niveau qui transcende celui de l'individu. Effectivement, si l'on considère le langage et son rôle dans *la transmission culturelle* pour décrire les concepts, on s'aperçoit que les régularités de ces derniers ne reposent pas tant sur des contraintes internes et innées fortes, et surtout, on s'aperçoit que les concepts peuvent être créés, accumulés et enrichis dans le temps et à travers les générations dans une *structure interindividuelle*. Bref, les véritables concepts sont peut-être le fruit d'un travail collectif et linguistique, alors que les autres représentations généralisantes de type1 (type1) sont peut-être plus facilement explicables par des mécanismes idiosyncratiques.

2-Entre Frege et le psychologisme

En fait, il n'est pas si étrange de penser aux contenus des concepts comme des choses qui transcendent des crânes. Par exemple, au 19^{ième} siècle alors que le courant *psychologiste* soutenait que les mathématiques, la logique, les théories de la pensée et de la signification renvoyaient à la conscience subjective (internes) Frege,

lui, (et d'autre: Bolzano, Husserl) a « expulsé les pensées hors de la conscience »[124]. Frege a ainsi séparé de manière radicale le subjectif de l'objectif : les pensées devenaient objectives puisqu'elles étaient publiques et accessibles par tous, tandis que les « représentations mentales » étaient subjectives et privées puisqu'elles étaient relatives à une personne.

Aujourd'hui, la division entre subjectif et objectif n'est plus aussi tranchée et le relativisme n'est plus autant à craindre. Une représentation interne n'est pas nécessairement une chose purement subjective ou privée. Au contraire, en sciences cognitives, on peut s'attarder autant aux propriétés générales des représentations mentales qu'à leurs contenus. De la sorte, sans tenir compte de leurs contenus, elles peuvent aujourd'hui être comprises comme des espèces naturelles (exemple, prototype, théorie, idéal, etc.) et puisque les mécanismes d'acquisition propres à ses représentations sont universels, les différents individus peuvent partager des contenus semblables et ainsi espérer communiquer des contenus[125] (même s'il n'en demeure pas moins qu'un individu seul ne reconstruit pas tous les concepts qui existent déjà. Ni un enfant, ni un robot ne peuvent s'éduquer eux-mêmes en confrontant leurs mécanismes représentationnels et leur environnement.)

D'autre part, si l'internalisme n'est plus à craindre pour son relativisme, l'externalisme ne l'est plus pour son ambiguïté ontologique. C'est-à-dire que l'idée des contenus publics et externes de Frege ou même d'un troisième lieu entre la référence et la représentation interne est moins à craindre que ce que l'étiquette *néo-platonisme* ne le suggère. En effet, si l'expulsion des concepts dans un « troisième monde » où ils subsisteraient indépendamment de deux autres mondes est le moyen pour Frege de les objectiver, pour nous cette expulsion des concepts dans un espace abstrait nous permettrait davantage d'expliquer leur création depuis le moment zéro,

---

[124] Dummett. M. *Les origines de la philosophie analytique*. Paris, Gallimard (p 37).

[125] Margolis, *Concepts core reading*, Edited by Eric Margolis and Stephen Laurence. 1999.

leur source. Partant, dans la mesure où l'extériorisation des concepts ne sert plus à préserver leur objectivité, mais à expliquer la source de leur contenu, nous ne sommes pas obligés d'invoquer un espace complètement détaché ontologiquement de l'activité cognitive. Ce n'est pas nécessaire de faire appel à un monde intelligible ou même à une *structure logique* entre les pensées, le langage et le monde[126]. L'idée de Frege d'un espace public peut, de la sorte, nous aider à penser *en dehors de la boîte* internaliste et à élucider certaines questions comme celles qui consistent à expliquer physiquement comment des concepts existent dans un *espace abstrait qui sert d'orienteur aux apprenants* avant qu'un sujet ne les acquière (ch2). Or, le lieu à l'extérieur des consciences individuelles mais *sur terre* où peuvent loger les concepts c'est le langage. Le langage comme banque de concepts intersubjective et intergénérationnelle peut nous servir d'espace public matériel qui transcende l'individu et qui répond aux problèmes de l'origine des contenus et de la supervision de l'apprentissage.

En effet, si le langage est défini comme un espace commun distribué où sont stockés les concepts et où il est possible de les modifier, retravailler, saboter, raffiner de manière collaborative, alors il est possible d'expliquer comment il remplit la fonction de cible pour orienter la cognition individuelle et surtout comment ils sont créés depuis leur origine, et ce, tout en respectant des postulats naturalistes et matérialistes. Dans ce sens, nous verrons plus en détails comment l'étude du langage sous cette perspective permet de comprendre l'assimilation culturelle d'innovations cognitives individuelles à un moment particulier : l'exploitation et la préservation de ces innovations; la transmission de ces innovations aux générations suivantes; et la régulation toujours plus sophistiquée de la cognition individuelle. Bref, il est possible de soutenir que le langage est l'espace externe qui peut nous permettre d'expliquer l'histoire des concepts et donc leur véritable origine.

---

[126] Wittgenstein,

Certains soulèveront ici le fait que ces processus collectifs ne sont pas ignorés de la tradition philosophique, avec raison. Plusieurs écoles se disputent pour expliquer les processus collectifs relatifs au langage qui expliquent les transformations conceptuelles. Le séjour de l'Esprit dans la conscience individuelle comme Hegel le développe. La continuité de l'évolution par la sélection naturelle au niveau des mèmes comme le pense Dawkins et Dennett. La marche de la science vers l'Ultime Théorie Vraie comme l'espère l'empirisme logique. La compétition entre différentes institutions pour l'espace médiatique et l'argent comme le pensait Khun. Même Churchland dans un texte plus récent avoue qu'il faut fournir une telle explication de l'évolution collective des représentations.[127] Il ajoute même que ce qui importe c'est qu'il faut expliquer comment les mécanismes culturels régularisent et accroissent la cognition individuelle. Toutefois, il n'avance aucune réponse.

D'autres comme W. Prinz, Dennett et Donald, par exemple, semblent fournir de telles explications. Ils réussissent à montrer la nouveauté que le langage apporte et le degré avec lequel il lance l'humanité sur une trajectoire intellectuelle impossible à toutes autres créatures[128] (ch3). Toutefois, personne parmi ces auteurs canons ne propose à ce jour de descriptions ni d'explications simples et concrètes, selon nous, des mécanismes cognitifs collectifs relatifs au langage qui créent les concepts depuis zéro. Il y a toutefois une exception, c'est celle d'Hutchins et d'Hazelhurst qui avancent un modèle d'architecture cognitif distribué.

En fait, il manque à la grande majorité des auteurs une définition du langage comme médium public qui incarne physiquement la compréhension conceptuelle des individus. En effet, nous comprenons le langage comme une cible sur laquelle les mécanismes de conceptualisation individuel des membres d'une collectivité doivent

---

[127] Churchland, Paul, *OUTER SPACE AND INNER SPACE: THE NEW EPISTEMOLOGY*, ED.:APA, UNIVERSITY OF CALIFORNIA, SAN DIEGO, 2002

[128] Voir par exemple D. Dennett, *From Animal to Person : How Culture Makes Up our Minds*; ou Marlin W. Donald, *Cognitive Evolution and the Definition of Human Nature,*

se conformer. Les jeunes des nouvelles générations ne commencent pas à conceptualiser à partir de zéro, mais à partir de ce que leur ont légué la génération précédente à travers le langage. Les jeunes apprenants et utilisateurs du langage sont les héritiers des réalisations cognitives individuelles de leurs ancêtres. En apprenant le langage de leurs parents et de leur société, ils orientent le développement de leurs concepts en conformité avec une structure conceptuelle qui a déjà été testée par une communauté d'agents cognitifs qui les précèdent.

À partir d'ici, les processus de construction de concepts ne se réduisent plus à la cognition individuelle produite dans le laps de temps correspondant à une vie humaine. L'espace abstrait, extérieur et collectif contenant les représentations incarne des myriades d'innovations cognitives effectuées par des individus particuliers et accumule ces innovations sur des centaines de générations. De plus, cette banque de concepts que le langage implémente peut évoluer à travers la succession des générations pour devenir plus sophistiquée, nuancée, proche de la réalité que la banque de concepts en vigueur dans les générations passées. Bref, une macro histoire naturelle de la conceptualisation est possible.

En résumé, l'argumentation précédente laisse planer l'idée qu'il est peut-être possible de comprendre les représentations de type1 comme le produit d'un travail individuel, alors que les représentations de type2 seraient plutôt le fruit d'une activité collective. Est-ce que le caractère collectif des représentations serait le véritable critère pour définir les concepts? C'est-à-dire pour les distinguer des autres types de représentations généralisantes?

Pour s'en convaincre et demeurer cohérent avec les postulats des sciences cognitives, il faut regarder au plan de l'architecture des agents cognitifs. Est-ce qu'il est possible de penser un modèle d'architecture cognitive qui puisse à la fois réaliser des représentations ou des complexes de représentations de type 1 et produire en

parallèle des représentations linguistiques, voire des concepts, qui entreraient dans un commerce collectif et intergénérationnel? Dans le même sens, est-il possible de modéliser une communauté d'agents cognitifs qui créerait des concepts à partir de zéro à l'aide de ce commerce public de représentations?

Ce qui nous manque pour compléter les modèles de RN à la Churchland tels que décrits dans le chapitre deux, c'est un deuxième niveau linguistique qui permettrait à plusieurs RN de s'échanger de l'information sur le monde. Il manque une architecture individuelle qui peut traiter des représentations linguistiques sur un plan et continuer à traiter l'information perceptuelle en même temps. Par exemple, il faudrait qu'un RN puisse générer des prototypes de manière individuelle et produire ensuite des symboles à partir de ces représentations, lesquelles seraient partagées à leur tour dans un espace public externe.

3- Communauté de RN comme unité d'analyse

Dans ce sens, Hutchins et Hazlehurst soutiennent qu'à partir du moment où une représentation existe, par exemple dans la tête du superviseur en laboratoire, il n'est plus difficile d'expliquer comment elle peut être transmise à un réseau de neurones. Ce qui est plus difficile, c'est d'expliquer comment peut émerger une représentation dans une population d'individus sans que personne de l'extérieur ne la donne. Comment un concept peut naitre sans qu'il n'existe déjà dans un commerce cognitif formant un espace abstrait?

Pour l'expliquer, Hutchins et Hazlehurst ont conçu des modèles connexionnistes qui considèrent les représentations comme un produit non seulement de l'interaction entre un individu et son environnement, mais aussi de l'interaction entre plusieurs individus d'une population. L'intention de départ est de prendre les aspects sociaux

et cognitifs sur le même piédestal, c'est-à-dire de prendre une communauté d'esprits comme unité d'analyse au lieu de prendre l'individu isolément. Avec cette perspective, de nouvelles propriétés apparaissent dans la théorie. Ça permet l'étude du rôle de l'organisation sociale sur l'architecture cognitive du système et permet de décrire les conséquences cognitives de l'organisation sociale à un niveau communautaire. Ça permet finalement aux phénomènes représentationnels qui sont en dehors des individus, ce que Hutchins et Hazlehurst nomment «artéfacts», d'être traités comme de vraies parties de l'analyse cognitive[129].

Dans *How to built a lexicon*[130], Hutchins et Hazlehurst montrent qu'en prenant une population de RN sur plusieurs générations d'individus comme niveau d'études, il est possible de simuler la construction de représentations linguistiques. Pour penser ce processus inter-individuel ou culturel, l'on doit considérer de manière concrète et matérielle le partage de représentations entre des individus. Chez l'humain, on utilise des systèmes de symboles, des langages, pour partager des représentions. Un symbole est une représentation constituée d'une *forme* (un mot par exemple) et d'un *sens* (sa signification). Dans leur article, Hutchins et Hazlehurst présentent un modèle qui simule comment un lexique de symboles partagé dans une communauté de RN peut émerger de l'interaction de ces simples agents cognitifs.

Ce modèle ne prétend pas expliquer tout le traitement des représentations chez l'homme, mais montre comment une structure représentationnelle symbolique peut émerger là où il n'y en avait pas auparavant: comment un lexique peut émerger dans une communauté d'agents cognitifs.

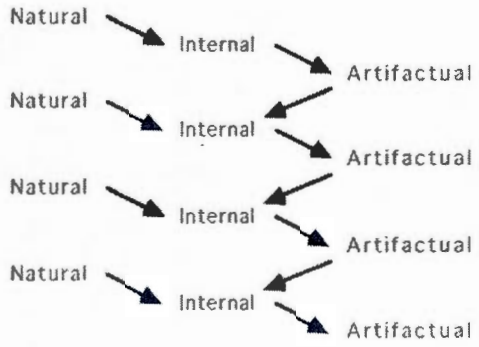
---

[129] Edwin Hutchins et Brian Hazlehurst, *Learning in the Cultural Process*, Department of Cognitive ScienceUniversity of California, 1990

[130] Edwin Hutchins et Brian Hazlehurst, *How to invent a lexicon: the development of shared symbols in interaction*, Department of Cognitive Science University of California,

Chose intéressante pour nous, ce processus culturel modélisé par Hutchins et Hazlehurst ne se limite pas aux structures mnémoniques: en plus des *structures internes* dans le crâne des individus, il y a des *structures naturelles* dans l'environnement et surtout des *structures artéfactuelles* qui sont échangées entre les individus dans l'environnement. Mises en interaction, les structures artéfactuelles forment un pont entre celles internes (dans un même individu ou entre plusieurs), c'est-à-dire qu'elles médiatisent un commerce entre les représentions internes de plusieurs individus sur plusieurs générations; à leur tour les structures internes forment un carrefour pour différentes structures artéfactuelles, en plus de former le seul pont possible entre les structures naturelles et celles artéfactuelles.

Figure 4.1: Trois structures 1



Plus concrètement, il y a trois types d'apprentissage opérés par les individus pour que la dynamique communautaire fonctionne: primo, un apprentissage direct des régularités naturelles de l'environnement, deuzio, un apprentissage médiatisé des régularités naturelles par les descriptions qui sont dans les structures artéfactuelles et tertio, un apprentissage d'un langage qui permet une cartographie entre les structures des régularités naturelles et les descriptions qui sont dans les structures artéfactuelles.

Le but de ce modèle est de montrer que si chaque individu est capable d'apprendre quelque chose à propos des régularités naturelles de l'environnement et ensuite de représenter ce qu'il a appris sous une forme (artefactuelle) qui peut être utilisée par d'autres individus pour faciliter leur apprentissage, alors le savoir à propos des régularités de l'environnement peut être accumulé dans le temps et à travers les générations. Par conséquent, les populations peuvent découvrir des choses impossibles à découvrir dans la seule vie d'un individu. Bref, pour décrire les représentations ici symboliques dans toute leur richesse, il faudra dépasser la perspective psychologique et prendre en considération les structures artéfactuelles qui seules peuvent expliquer ce phénomène de rétention des connaissances entre les générations.

3.1- Simulation du modèle de Hutchins et Hazlehurst

Comment techniquement Hutchins et Hazlehurst simulent-ils ce processus culturel? D'abord Hutchins et Hazlehurst utilisent des simulations de RN appelées «autoassociateurs», c'est-à-dire des RN qui utilisent, comme la plupart des autres, une procédure de *rétropropagation de l'erreur* pour trouver la configuration de poids synaptiques à atteindre[131], toutefois avec les «autoassociateurs» la sortie n'est pas analysée par le chercheur, mais elle est simplement comparée au vecteur de la couche d'entrée (dont la configuration est causée par un stimulus extérieur). C'est-à-dire que la cible qui oriente l'apprentissage ne provient plus directement du chercheur, mais d'une structure naturelle imprégnée sur le capteur[132].

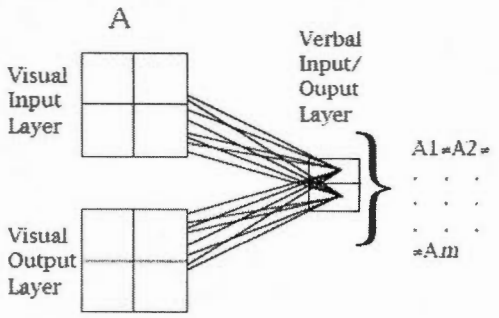
Maintenant, Hutchins et Hazlehurst proposent de *plier* en deux leur RN à trois couches, ce qui permet à la couche cachée de produire une représentation qui peut devenir une partie matérielle importante de l'interaction inter-individuelle. C'est-à-

[131] Voir chapitre 2

[132] Théoriquement, puisque l'environnement est entièrement contrôlé par l'expérimentateur.

dire que la couche cachée devient publique dans la mesure où elle produit des représentations qui remplissent une fonction importante d'un lexique: c'est par cette couche que leur agent produit des mots et en reçoit pour les comparer avec les mots des autres agents. Hutchins et Hazlehurst rebaptisent la couche cachée «Entrée/sortie-verbale». Par exemple, si les autres couches du réseau imitent la fonction d'un système visuel très simple, capable de classer des scènes dans l'environnement, alors la couche «d'entrée/sortie-verbale» est capable de générer des schémas d'activation en réponse à chaque scène visuelle rencontrée.

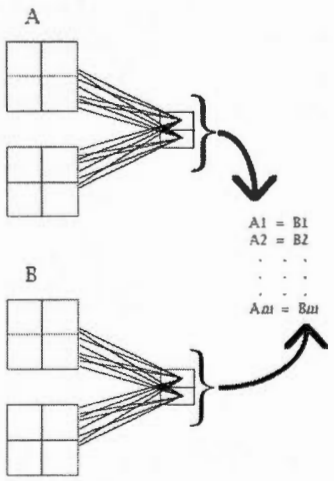
Figure 4.2: Réseau de neurones 1



Ensuite les RN doivent pouvoir partager leurs mots et leurs significations. Rappelons d'abord que traditionnellement les RN sont utilisés pour modéliser des propriétés de la cognition propre à des individus. Hutchins et Hazlehurst, eux, suggèrent d'utiliser une population de RN pour modéliser les propriétés d'une communauté de réseaux. Dans le modèle classique, celui de Churchland par exemple, le monde appris par un RN est donné ou construit par le programmeur. C'est-à-dire que dans les modèles classiques, c'est la valeur donnée (patron de configuration cible) à la couche de sortie qui oriente l'apprentissage. Le modèle communautaire suggère plutôt que le comportement des membres d'une population soit aussi une source importante de structure à partir de laquelle chacun peut apprendre. Une grande part de la

supervision de l'apprentissage d'un RN est effectuée par d'autres RN grâce à l'espace où transigent les structures artéfactuelles. Techniquement, Hutchins et Hazlehurst montrent que deux RN peuvent facilement arriver à un consensus en permettant aux «entrées/sorties-verbales» (de la couche médiane) de l'un de devenir la cible qui supervise l'apprentissage de l'autre. Si chacun prend la «sortie-verbale» de l'autre comme cible le consensus survient.

Figure 4.3: Communication entre deux RN 1



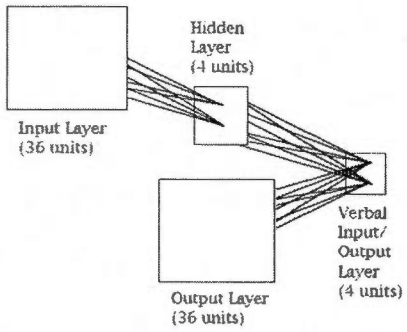
Finalement, dans leurs simulations, Hutchins et Hazlehurst montrent qu'une population de RN arrive à créer et partager un ensemble de patrons d'activités sur les unités d'«entrées/sorties-verbales» qui font la distinction entre une multitude de scènes naturelles. Ces lexiques de patrons (mots) sont des consensus concernant le bon patron d'activation à attribuer à une scène naturelle qui apparait suite à des milliers d'observations individuelles mises en commun. Ces lexiques ne sont pas préétablis par le superviseur, au contraire, ils ne sont même pas prédictibles. C'est vraiment le commerce entre les milliers de perceptions internes, toutes différentes au départ, qui servent «d'entrée-verbale» ou de cible pour les autres qui permettent la création d'un consensus. Par la suite, un individu qui nait dans une population où il y a consensus apprend en écoutant les autres à attribuer le bon patron d'activation à la

bonne scène naturelle. Cet individu peut maintenant participer à raffiner les représentations publiques du monde en continuant de confronter les patrons issus du consensus culturel (inter-générationnel) à ses observations. Ainsi, les populations peuvent découvrir des choses, construire des représentations, impossibles à découvrir dans la seule vie d'un individu.

Par exemple, une des simulations que proposent Hutchins et Hazlehurst met en scène des réseaux autoassociateurs à quatre couches dans un environnement où une lune se meut suivant un cycle d'une douzaine de phases. Les RN possèdent tous une couche d'entrée *visuelle* à 36 unités, une couche cachée de 4 unités, une couche intermédiaire d'entrée/sortie verbale de 4 unités et finalement, une couche de sortie *visuelle* de 36 unités[133]. Lorsqu'un RN se fait présenter un événement particulier de l'environnement, à savoir une phase de la lune, il produit d'une part une sortie verbale, un symbole ouvert à la communauté et, d'autre part, il produit plus en amont une sortie finale qui a pour cible l'entrée de la première couche.

Figure 4.4: RN autoassociateur 1

[133] Notez que la deuxième couche intermédiaire (additionnelle) facilite la compression des données à l'entrée visuelle et permet aussi d'effectuer le travail relatif aux valeurs d'activations de l'entrée verbale.



Maintenant, 5 individus entrent en communication entre eux de façon randomisée. Précisément, deux à la fois, un *orateur* et un *auditeur* se font présenter une scène, une phase de la lune, pour laquelle l'orateur produit un patron d'activation sur sa couche de sotie verbale. Le deuxième individu produit aussi une représentation de ce qu'il dirait dans ce contexte, mais comme auditeur, il se sert du patron d'activation verbale de l'orateur comme cible pour corriger, par rétropropagation, ses propres erreurs. De plus, l'auditeur et l'orateur opèrent une époque d'apprentissage classique puisque leur propre sortie verbale sert d'entrée à leur couche de sortie qui produit un patron d'activation *visuel* se rapprochant ultimement du stimulus premier: la phase de la lune elle-même. En bout de ligne, cette architecture permet d'une part que la sortie *verbale* de l'auditeur ressemble davantage à celle de l'orateur et d'autre part que la sortie *visuelle* du RN ressemble davantage à la scène dans le monde.

L'objectif est que chacun des 5 membres communiquent avec tous les autres et arrivent à un consensus à propos des représentations produites à partir des 12 phases de la lune. Au départ, la configuration des poids de connexions des individus est établie de manière randomisée. Partant, les représentations linguistiques produites par la couche d'entrée/sortie verbale ne discriminent pas entre les phases de la lune.

Après 2000 interactions verbales entre chacun des RN, tous répondent différemment devant chacune des 12 phases de la lune, par contre tous s'entendent sur la manière de répondre. Il y a consensus sur la manière de discriminer les événements.

Fait important, puisque les poids de connexions sont randomisés et que le protocole d'interaction qui organise les expériences de chaque RN l'est tout autant, il est impossible de prédire quel lexique les RN développeront. Toutefois, un lexique quelconque émerge presque toujours de la procédure.

Soit, cet exemple montre comment, matériellement, un lexique peut apparaitre dans une communauté d'individus, toutefois le lecteur ne sera peut-être pas encore convaincu de la nécessité ou du moins de l'immense avantage d'être en communauté pour apprendre à propos des régularités et/ou des invariants dans le monde. Pour cette raison, une deuxième simulation appuie notre thèse, à savoir qu'il faut être plusieurs pour produire la représentation d'un invariant, voire d'un concept, depuis son origine.

3.2- Simulation de l'apprentissage dans un processus cultuel

Il y a plusieurs centaines d'années, des tribus d'autochtones pêchaient les fruits de mer à marée très basse sur le site de l'Université de Californie à San-Diego. Puisque ces tribus ne vivaient pas près des sites de pêche, il était rentable de se déplacer uniquement lorsque les eaux se retiraient suffisamment pour permettre la cueillette. Un individu qui ne possédait pas la connaissance sur les marées pouvait passer des semaines à en attendre une assez basse. De plus, il était impossible de prédire si la marée deviendrait assez basse à une occurrence particulière seulement en regardant un court moment. Elle pouvait très bien s'avérer forte, moyenne ou faible.

En fait, ces tribus avaient appris à prédire le cycle des marées en utilisant les phases de la lune comme facteur prédictif. Lorsque la lune est pleine ou nouvelle, cette dernière est en phase de produire une grande variation de marées (en réalité, le soleil joue aussi un rôle important dans la marée), c'est-à-dire, qu'une très haute et une très basse marées se produisent le même jour.

Maintenant, Hutchins et Hazlehurst proposent d'imaginer une tribu qui possède les concepts de marée et de phases lunaires, mais qui ne conceptualise pas encore la corrélation entre les deux. Dès lors que ça prend de très longues heures pour observer le cycle des marées et que le ciel n'est pas toujours dégagé pour faire une corrélation avec les phases lunaires, il n'y a que quelques opportunités pour un individu de conceptualiser cette association, sûrement trop peu selon Hutchins et Hazlehurst, pour que cet individu seul se représente cette régularité. Toutefois, on peut imaginer une communauté qui réussisse à résoudre, à travers le temps et les générations, ce type de problème trop complexe pour un individu seul. Imaginons une communauté en particulier qui possèdent déjà un langage développé, parmi laquelle chacun des membres apprend dans son développement normal un ensemble partagé de relations représentationnelles entre, d'une part, les phases de la lune et les mots pour nommer les phases de la lune; et d'autre part, entre les états des marées et les mots pour ces états. En fait, imaginons une communauté où il s'est développé un consensus à propos de lexiques, comme celui exposé dans la section précédente. Rappelons toutefois que dans leur simulation, Hutchins et Hazlehurst, s'ils assument que ces concepts ou représentations sont bien connus des individus, la conjonction entre les phases de la lune et les états de la marée, elle, demeure à conceptualiser.

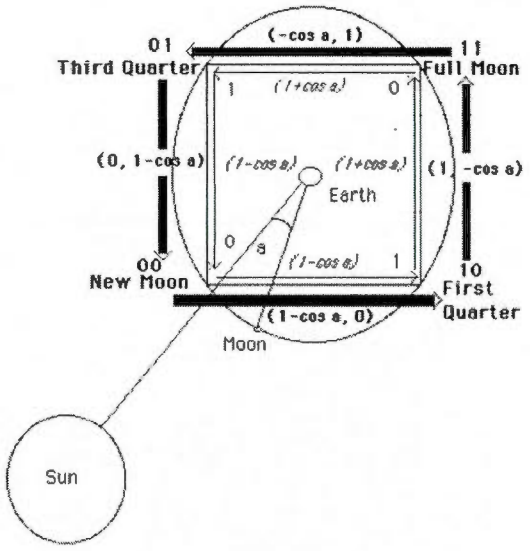
En termes plus techniques, dans la simulation de Hutchins et Hazlehurst, la population est constituée de RN classiques. Ces derniers sont capables de recevoir des entrées tant d'un environnement naturel que d'un autre artéfactuel (culturel).

Durant leur vie, chaque RN a la chance de se confronter à la tâche qui est de corréler les phases de la lune aux marées et de produire le bon artéfact qui représente l'invariant (concept publiable). Après son expérience, le RN produit un artéfact, donne naissance à un rejeton et meurt. Ensuite, les rejetons répètent l'expérience sans rien hériter de façon génétique mis à part le fait d'être un RN configuré de manière randomisée. De la sorte, la seule contribution innovante qu'un individu peut offrir à ses successeurs est culturelle.

Il y a 28 paires de phases-lune/états-marée qui constituent de l'information sensorielle directe à propos de l'environnement. Ces 28 paires sont générées par la division en 28 de l'orbite lunaire et de l'encodage des phases de la lune et des états de la marée pour chacun de ces jours du mois lunaire.

Maintenant, chacun des 2 éléments du vecteur qui représente les phases de la lune forme un chiffre entre 1 et 0. Les quatre combinaisons possibles, 00, 10, 11, 01, représentent les quatre phases majeures de la lune : nouvelle, premier quart, pleine, trois quarts. Ces quatre phases majeures sont représentées par les quatre coins d'un carré (voir figure 4.5). Notez que le premier élément du vecteur encode la partie gauche de la lune qui est visible depuis une Terre idéalisée, alors que le deuxième élément du même vecteur encode la partie visible droite.

Figure 4.5: représentation de l'environn 1



Les états de la marée, eux, sont encodés par un seul élément, un chiffre entre 1 et 0 généré par un calcul de l'angle que forment la lune et le soleil par rapport à la Terre. Ainsi, chaque côté du carré représentant les phases de la lune est associé avec une variance de la marée soit croissante soit décroissante.

Les membres de la communauté doivent apprendre à dire si la marée est propice ou non à partir d'une observation de la lune. Pour ce faire, ils développent deux petits lexiques qui représentent deux classes d'événements: les phases de la lune et les états de la marée. Nous verrons plus précisément plus bas que trois (sous)réseaux de trois couches forment un RN membre de la communauté. Un (sous)réseau encode en entrée les quatre phases de la lune, un deuxième (sous)réseau encode en entrée les états de la marée et un troisième (sous)réseau reçoit en entrée les sorties verbalisées ou symbolisées des deux autres (sous)réseaux pour résoudre la tâche finale qui consiste à corréler les phases de la lune et les états de la marée.

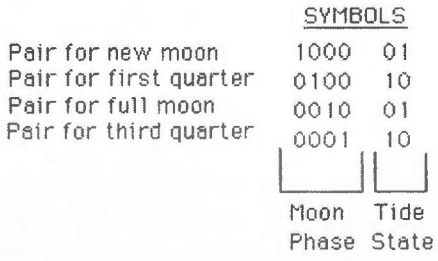
Figure 4.6: représentation du langage 1

ENVIRONMENT	SYMBOLIC REP	PHYSICAL REP
	Prototypic Moon Lexicon	Moon Phase
New moon	"1000"	00
First quarter	"0100"	10
Full moon	"0010"	11
Third quarter	"0001"	01
	Prototypic Tide Lexicon	Tide State
Large-variance tide	"01"	0
Small-variance tide	"10"	1

TABLE 1. Citizen Language.

De plus, dans la simulation de Hutchins et Hazlehurst, des artéfacts publics et en ce sens externes représentent la corrélation entre la lune et les marées (la tâche ultime du RN) en agençant 4 paires de symboles. Le premier symbole de chaque paire est un symbole propre à une des 4 phases majeures de la lune et le deuxième symbole est propre à un état de la marée.

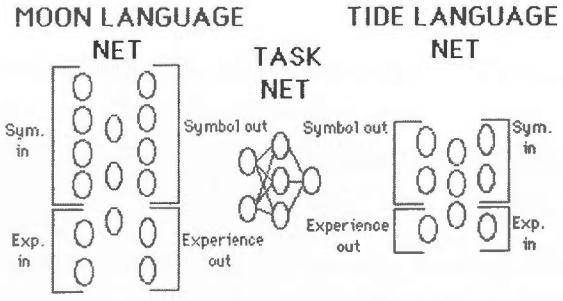
Figure 4.7: artéfacts 1



Comme mentionné ci-dessus, un RN est composé de trois (sous)réseaux de neurones *feedforward :* deux (sous)réseaux de «langage» et un destiné à la «tâche». Maintenant, chacun des (sous)réseaux est autoassociateur, à savoir qu'il est entrainé à reproduire sur la couche de sortie ce qui est appliqué sur la couche d'entrée. Or, sur la couche d'entrée, les (sous)réseaux de «langage» reçoivent du même coup: une

description symbolique d'un évènement et l'évènement lui-même. En d'autres termes, le RN fait l'expérience, individuellement, d'un événement, en l'occurrence de l'état de la marée ou de la phase de la lune, et simultanément, il se fait décrire la scène par un compatriote (via un artéfact). En bout ligne, chaque (sous)réseau de «langage» est entrainé à opérer la concaténation des deux types d'information (symbolique et physique) et ainsi à produire une représentation qui conjugue les deux. Plus précisément, le (sous)réseau est capable de générer une représentation symbolique à partir de l'expérience d'un évènement et de faire simultanément l'expérience d'un évènement à partir d'une représentation symbolique acquise.

Figure 4.8: Architecture d'un RN citoyen 1



Partant, il y a 3 types d'apprentissage propres à ces RN. Le premier type est l'opération du (sous)réseau de langage que l'on vient de décrire qui consiste à

associer les symboles et les événements en des patrons uniques de sortie qui, à leur tour, servent d'entrée au (sous)réseau de la «tâche».

Deuxièmement, ce dernier réseau du centre qui doit opérer la «tâche» principale de tout le RN doit d'abord apprendre à associer les représentations des phases de la lune avec celles des états de la marée. En l'occurrence, le chercheur offre une série limitée d'expériences des deux événements de manière simultanée pour qu'il puisse ainsi apprendre à prédire l'un avec l'autre. L'objectif est que, lorsque l'on choisit une journée parmi les 28 du cycle lunaire et que l'on présente une représentation de la phase de la lune correspondant à cette journée comme entrée au réseau, ce dernier émet une prédiction de l'état de la marée en sortie. Celle-ci est comparée à l'état réel qui correspond à cette journée. Si une erreur est détectée, elle est *rétropropagée* pour ajuster la configuration de poids de connexions. Cette apprentissage ne nécessite pas la faculté de langage nous disent les auteurs, mais est produite par la présentation d'une entrée et d'une cible directement au (sous)réseau de la «tâche», sans passer par les deux autres propres aux «langages».

C'est plutôt lors du troisième type d'apprentissage, *l'apprentissage médiatisé*, que la faculté du «langage» est utilisée. Ce type est caractérisé par l'utilisation des 2 (sous)réseaux de «langage» qui produisent des données en sortie, lesquelles servent d'entrées et de cibles pour le (sous)réseau propre à la «tâche».

Figure 4.9: apprentissage médiatisé 1

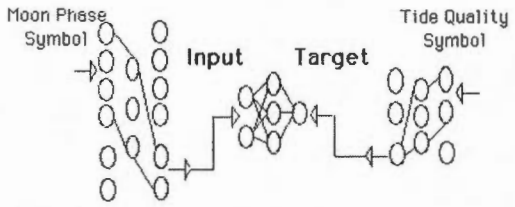


FIGURE 6. Mediated Learning. The language provides interpretations of inputs and targets for task learning.

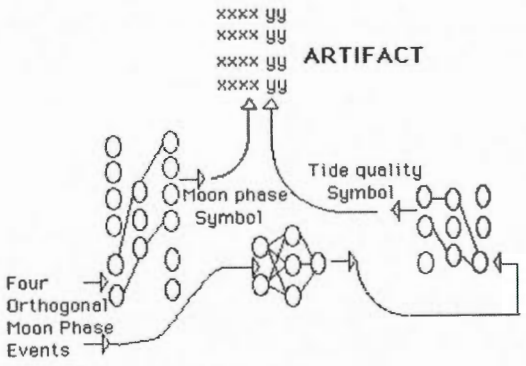
Maintenant, le fait de passer par les 2 (sous)réseaux du «langage» met en interaction, d'abord à l'intérieur de l'individu, une description symbolique et une expérience individuelle. Ensuite, à l'extérieur de l'individu, des artéfacts (des résultats symbolisés d'expériences individuelles qui sont déposés par d'autres membres dans un espace public) rendent possible l'interaction entre les multiples événements dans le monde et les informations à leur propos. En d'autres termes, les artéfacts repassent constamment, suivant les générations, devant le *tribunal de l'expérience* à travers les RN individuels. De la sorte, ils sont constamment transformés, raffinés, édulcorés, etc., en fonction du fil des événements qui se présente à la communauté de RN.

Maintenant, pour demeurer fidèle à leur thèse externaliste, Hutchins et Hazlehurst ont voulu que l'apprentissage de la tâche soit impossible à réaliser par un individu seul, mais seulement par une communauté à travers des générations. Toutefois, ils ont voulu aussi que l'apprentissage soit possible par un seul individu une fois qu'un bon artéfact a été développé par la communauté. Pour ce faire, chaque individu est limité à 750 époques d'entrainement sur un artéfact (puisque chaque artéfact a 4 instances d'apprentissage cela fait 3000 procédures). De plus, chaque individu reçoit 260 épreuves d'apprentissage à partir d'expériences directes avec l'environnement. Avec ce protocole d'apprentissage, la chance qu'a un individu de produire le bon artéfact seul est proche de zéro, alors qu'il est possible que la culture (plusieurs RN sur plusieurs générations qui partagent des artéfacts) en soit capable. Et si la culture peut produire un artéfact capable de représenter les régularités du

monde, la combinaison entre quelques époques d'apprentissage direct avec l'environnement et beaucoup d'époques d'apprentissage médiatisé par des artéfacts devrait, comme mentionné ci-dessus, permettre à un individu seul d'apprendre à prédire correctement les régularités.

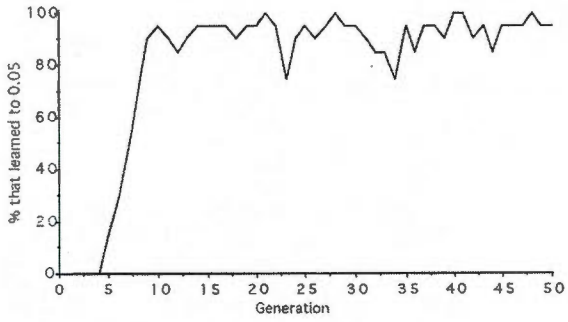
Maintenant, après avoir appris, à sa mesure, sur la régularité qui existe entre les phases de la lune et les états de la marée, un RN est invité à produire un artéfact en «répondant» à un petit test de connaissances. Précisément, on soumet en entrée les quatre phases de la lune (pas les symboles, mais bien les événements : 00, 10, 11, 01) à partir desquels le (sous)réseau de «langage» entrainé pour observer la lune produit des symboles en sortie et, à partir desquels le (sous)réseau de la «tâche» fournit les représentations des états de la marée correspondant à chacune des phases de la lune en sortie. Ensuite, l'autre (sous)réseau de «langage» entrainé pour verbaliser les états de la marée reçoit en entrée les représentations produites par le (sous)réseau de la «tâche» et produit à son tour des symboles des états de la marée. Si le réseau a bien appris, les deux groupes de symboles, ceux qui réfèrent aux phases de la lune et ceux qui réfèrent aux états de la marée, devraient représenter la régularité *derrière*.

Figure 4.10: production d'un artéfact 1



Un des résultats intéressant pour nous est qu'au début de la simulation, lorsque qu'une communauté de 20 RN est placée face à la tâche, aucun individu n'est capable seul de prédire la régularité dans l'environnement. La structure artéfactuelle est trop pauvre pour superviser l'apprentissage. Mais après quelques générations de RN qui participent à médiatiser les artéfacts publics et les événements extérieurs dans leur espace interne (à tenter de les faire correspondre) une structure artéfactuelle représentant adéquatement la réalité se met en place. Partant, en utilisant le même protocole d'apprentissage et la même architecture cognitive, presque tous les individus des dernières générations sont capables d'apprendre la tâche, à savoir de prédire l'état de la marée à partir de l'observation des phases de la lune. Ce résultat montre bien qu'il n'est pas nécessaire d'observer une évolution des capacités innées d'apprentissage pour observer une évolution dans la richesse du contenu appris. Ce phénomène est plutôt redevable à la rétention des représentations dans les artéfacts publics à travers les générations.

Figure 4.11: RN capable d'apprendre 1



CONCLUSION

CONCEPTUALISATION: ENTRE LES STRUCTURES SOCIALES ET INDIVIDUELLES

La source d'inspiration de ce mémoire est la difficulté de conjuguer le caractère interne ou psychologique des concepts et l'impossibilité qu'un individu, seul, puisse créer ces derniers. En effet, d'un côté les concepts sont traditionnellement définis comme des entités situées à l'intérieur d'agents cognitifs et comme servant de substitut formel à des réalités généralement extérieures. Plus précisément, nous avons montré qu'en sciences cognitives, on s'entend pour affirmer que les concepts sont des structures représentationnelles mises en réserve dans la mémoire à long terme et sont utilisés par différents processus de la cognition supérieure tels que la catégorisation, l'induction, la déduction, etc.[134] Ce sont en quelque sorte les *ensembles* de base de la connaissance commune à toutes ces opérations de la cognition supérieure. Par conséquent, définir les propriétés que partagent tous les concepts mémorisés est un préalable à une théorisation de la cognition supérieure en général. Maintenant, si tel est le cas, si on comprend les concepts ainsi, il n'est pas étonnant qu'on parte souvent d'observations sur l'utilisation de concepts par des individus pour les définir. C'est-à-dire qu'il est cohérent pour une théorie internaliste que les propriétés attribuées aux concepts s'accommodent de la façon qu'ont les

[134] Edouard Machery, *Concepts Are Not a Natural Kind*,

individus d'utiliser et d'acquérir des concepts lors de tâches expérimentales. Ils sont en quelque sorte une partie de l'organisme humain.

Or, d'un autre côté, nous avons montré au deuxième chapitre que si l'on tient compte de l'architecture cognitive physique (en plus de l'utilisation de contenus conceptuels) pour expliquer comment concrètement un individu ou un robot crée des concepts depuis leur origine, force est de croire que, seul, un agent tel que décrit généralement en sciences cognitives n'a pas les mécanismes pour le faire. Chez les humains par exemple, les chercheurs observent comment un individu apprend, reconstruit ou transmet un concept, mais très peu d'explications mécanistes apparaissent dans la littérature en sciences pour tenter d'expliquer comment un concept est créé depuis zéro dans l'univers par un individu. En fait, la question de l'origine d'un concept est souvent confondue avec celle qui consiste à savoir comment un concept *arrive* dans un individu. Par exemple, on tente moins d'expliquer, dans la perspective internaliste, comment le concept TABLE a été construit matériellement (dans le sens de créé) à son origine et transmis par la suite, que d'expliquer comment un enfant l'acquiert ou le (re)construit *dans sa tête.* On le constate aussi avec les réseaux de neurones qui sans supervision n'arrivent pas à créer de représentations généralisantes assez riches pour être subsumées sous le concept de concept.

En somme, dans la mesure où il est difficile de prouver l'inexistence ou l'impossibilité d'une chose, nous sommes obligés d'élaborer un argument par la négative. En l'occurrence, notre argument repose sur le fait que très peu d'auteurs s'intéressant aux concepts ne se soucient d'expliquer l'origine de ceux-ci alors même qu'ils prétendent les définir. Pourtant, en général, tant que la genèse d'une espèce n'est pas expliquée, la conception de sa nature repose sur une connaissance incomplète. Tant que la sélection naturelle comme mécanisme de l'évolution n'avait pas fourni une explication matérialiste et naturaliste à l'origine des espèces vivantes,

«la plupart des naturalistes croyaient que les espèces [étaient] des productions immuables créées séparément»[135]. Les auteurs contemporains ne partagent sûrement pas explicitement des croyances analogues en ce qui a trait aux contenus des concepts, mais ils font implicitement *comme si*. Dans le cas du concept de concept, tant qu'on ignore son mode de création depuis zéro, on ignore au minimum sa nature publique et externe et par conséquent, on est forcé de réduire sa nature à quelque chose d'essentiellement interne, voire de personnel[136]. Bref, en ignorant la question de l'origine des contenus des concepts, les auteurs qui s'engagent sur des définitions internalistes participent à maintenir un présupposé implicite (pas nécessairement voulu) à leurs thèses, à savoir que les contenus sont donnés d'emblée, et même des *productions immuables*.

Dans les deux derniers chapitres, nous avons montré d'abord que la solution est à chercher au niveau social. La production d'un concept est peut-être le fruit de l'interaction entre plusieurs individus d'une population : c'est-à-dire qu'un concept est possiblement construit à partir du commerce de représentations devenues publiques et externes. Plus concrètement encore, nous avons présenté les architectures cognitives de Hutchins et Hazlehurst qui nous ont permis de croire qu'il est possible de modéliser et de comprendre le travail de création du contenu des représentations davantage que les modèles plus classiques.

En effet, avec leur concept d'artéfact, Hutchins et Hazlehurst proposent d'abord une solution potentielle au problème de la supervision de l'apprentissage. Ces

---

[135] Charles Darwin, *L'Origine des espèces*. Éditeur, Paris, 1921 p.9 :
http://classiques.uqac.ca/classiques/darwin_charles_robert/origine_especes/darwin_origine_des_especes.pdf

[136] Le constructivisme radical, qui a influencé la réforme de l'éducation au Québec selon Normand Baillargeon, décrit les concepts comme des choses essentiellement personnelles. Voir, Normand Baillargeon, Le constructivisme radical ou comment bâtir une réforme de l'éducation sur du sable, in :
«Contre la réforme pédagogique», VLB éditeur. 2008.

représentations en étant publiques ou communicables remplissent la fonction de cible pour les apprenants. Du même coup, leur modèle fournit une explication matérialiste et naturaliste au problème de l'origine des représentations généralisantes dans la mesure où les artéfacts, par leur nature publique, permettent un travail incrémental qui dépasse les individus et les générations d'individus. Précisément, même si le premier artéfact créé dans une population (qui est fonction des mécanismes de catégorisation d'un seul RN) est loin de représenter une réalité dans le monde, l'amélioration progressive à coup de petites innovations cognitives fait en sorte qu'en bout de piste, l'artéfact évolue pour prendre la forme d'une représentation plus riche et complexe que celle produite par un seul individu. Par exemple, le concept de «marée basse corrélativement aux phases de la lune» possède un âge qui se calcule en dizaines de vies humaines distribué sur des centaines d'individus et assimile autant de petites innovations opérées par des agents cognitifs individuels.

Ainsi compris, un concept qui est reconstruit dans l'espace interne d'un individu possède des propriétés qui ne sont pas uniquement le fruit de mécanismes de catégorisation. Dans la simulation de Hutchin et Hazlehurst, même si chaque individu possède des (sous)réseaux de langage, à l'intérieur de cet individu, ces facultés linguistiques seules ne servent pas davantage que le (sous)réseau dédié à la «tâche». C'est-à-dire que, même si un RN possède une faculté linguistique, s'il est isolé, la production vectorielle de ce RN ne sera guère plus sophistiquée qu'un réseau classique. L'artéfact produit par un RN *sauvage* ou de première génération, qui n'a aucune propriété sociale, ne correspond pas plus à la régularité du monde qu'un vecteur produit par un unique (sous)réseau de «tâche» ou classique.

Cependant qu'en est-il des concepts? Est-ce que les représentations produites par les populations de neurones de Hutchins et Hazlehurst répondent aux critères qui circonscrivent le concept de concept? Essayons d'y répondre un critère à la fois.

Rappelons dans un premier temps, qu'à la différence des représentations généralisantes de type1, nous avons défini à la fin du premier chapitre le concept de concept comme un type de représentation qui participe aux facultés langagières. Ainsi ils peuvent prendre une forme linguistique.

Or, dans les simulations de Hutchins et Hazlehurst, si les représentations d'un appareil cognitif seul sont éminemment plus simples que les représentations produites par plusieurs de ces appareils cognitifs mis en collectivité, il faut croire alors, à l'instar des théories duelles, que la forme linguistique est un impératif mécanique pour qu'une représentation généralisante dépasse le niveau de sophistication qu'un individu seul puisse atteindre. En effet, les représentations qui émergent des collectivités sont nécessairement linguistiques puisque leur mode de production passe par la communication entre plusieurs individus. C'est précisément ce que Hutchins et Hazlehurst réussissent à montrer, à savoir que, lorsque les représentations prennent la forme de symboles linguistiques transmissibles entre les individus, il est possible d'expliquer la rétention d'informations qui transcendent les générations. Dans ce sens, Hutchins et Hazlehurst corroborent l'idée d'un critère linguistique qui distingue les représentations de type non linguistiques et celles de type linguistiques et dans le même sens, ils corroborent l'hypothèse que ce caractère linguistique confère aux représentations du deuxième type, un caractère externaliste ou supra-individuel.

Dans un deuxième temps, un des critères suggère que les concepts sont des représentations qui ne sont pas à propos de l'environnement directement, mais plutôt que leur rapport au monde passe par d'autres représentations plus simples. En effet, les concepts forment une synthèse de représentations singulières et, de la sorte, ils sont causés ou produits à partir de ces dernières et non directement à partir des stimuli de l'environnement. Ils sont médiatisés.

Or, le modèle Hutchins et Hazlehurst suggère l'idée que les représentations de type 2, linguistiques (donc social), sont en fait le produit d'un travail supplémentaire sur les représentations de type 1. Toutefois, contrairement à la tradition qui décrit ce travail comme celui d'une faculté supérieure individuelle (intellect, entendement, conceptualisation, etc.) sur des représentations singulières (des percepts), c'est plutôt une collectivité qui remplit ce rôle. C'est-à-dire que, lorsque le RN produit un artéfact ou un symbole public à partir de la concaténation de son expérience personnelle et de l'acquisition d'information d'autrui, mis à part la production de l'étiquette linguistique, il n'y a pas de véritable travail intellectuel plus sophistiqué que la catégorisation classique. En fait, le second travail est celui permis par la rétention de l'information dans un symbole linguistique et opéré par de multiples individus sur plusieurs générations. C'est ce travail collectif qui permet de saisir l'invariant ou la régularité derrière les perceptions toujours trop particulières et changeantes des individus. C'est aussi ce travail qui permet de produire des représentations, des artéfacts, qui ne sont pas le fruit d'une interaction entre un individu et son environnement, mais qui sont à propos de myriades d'autres petites représentations particulières et individuelles. Ce sont des représentations sur des représentations.

Dans le même sens, les représentations du modèle de Hutchins et Hazlehurst peuvent dans une certaine mesure remplir également le critère de la *référence différée* propre aux concepts, contrairement aux simples représentations de type1. En effet, puisque l'apprentissage peut être médiatisé, c'est-à-dire ici que, puisqu'un RN peut en apprendre sur le monde en se le faisant raconter, il peut reconstruire des représentations sans jamais en avoir fait l'expérience; ces représentations peuvent être à propos de choses qui dépassent l'horizon perceptible.

Soit, mais une fois qu'on a trouvé une solution partielle ou potentielle au problème de la supervision de l'apprentissage; qu'on a trouvé du même coup un modèle qui

fournit une explication de l'origine des représentations d'une certaine complexité; qu'on a commencé à satisfaire les critères *linguistique*, *médiatisé* et *différé* de la définition de concept, il nous reste à montrer comment dans ce système conceptuel (des RN en communauté) les individus procèdent de manière réfléchie et comment cette communauté peut produire des représentations qui prennent la forme de définitions. Malheureusement, l'espace qui nous était alloué étant épuisé, ces questions feront l'objet d'un autre travail.

BIBLIOGRAPHIE

Aristote, *Parva Naturalia*, extrait *in* recueil de texte: *philosophie de l'esprit et des sciences cognitives*. UQAM. 2004.

Aristote, *Éthique à Nicomaque,* CEC inc., 2010.

Aristote, *La Métaphysique* (trad. Pierron et Zévort), http://fr.wikisource.org.

Article *16. (1) du Code criminel canadien, http://laws-lois.justice.gc.ca/fra/lois/C-46/*

Bickle & Mandik, *The philosophy of neuroscience. In, Stanford Encyclopedia of Philosophy*. Site web

Cerveau à tous les niveaux: http://lecerveau.mcgill.ca/

Churcland Patricia, *Les neurosciences concernent-elles la philosophie*? *In*, Cahier de textes : *Philosophie de l'esprit et des sciences cognitives.* Prof. L. Faucher et P. Poirier et Beaulac. Montréal, presse de l'UQAM, aut. 2004.

Churchland Paul, *The engine of reason, the seat of the soul: a philosophical journey into brain.* First MIT Press paperback edition, 1996

Clark A., *Associative engine*. MIT Press. 1993

Dennett, D. C., *Heterophenomenology*, Tufts University, 2007

Dennett, D. C., *Freedom evolves*, Penguin Book, 2004

Dennett, D. C., *The Self as a Center of Narrative Gravity,* in F. Kessel, P. Cole and D. Johnson, eds, *Self and Consciousness: Multiple Perspectives,* Hillsdale, NJ: Erlbaum, 1992. Danish translation, *"Selvet som fortællingens tyngdepunkt,"* Philosophia, 15, 275-88, 19861986

Dumont, Louis, *Essais sur l'individualisme,* Seuil, 1983

Fodor and Lepore, *The Compositionality Papers,* Clarendon Press Oxford. 2002

Gardner, H. *The mind's new science: a history of the cognitive revolution.* New York: Basic Books, 1987

Hutchins et Hazlehurst, *How to invent a lexicon: the development of shared symbols in interaction*, Department of Cognitive ScienceUniversity of California,

Hutchins et Hazlehurst, *Learning in the Cultural Process*, Department of Cognitive ScienceUniversity of California, 1990

Kant, *Métaphysique des mœurs,* 1785. Traduction V. Delbos. Wikisource. 1907

Laurent, *Histoire de l'individualisme*, Presse universitaire de France, 1993.

Kirby, S., Dowman, M. and Griffiths, T. (2007). *Innateness and culture in the evolution of language*. Proceedings of the National Academy of Sciences

Machery, *Concepts Are Not a Natural Kind*, Depart. of History and Philosophy of Science, University of Pittsburgh. 200?

Margolis, *Concepts core reading,* Edited by Eric Margolis and Stephen Laurence 1999

McClelland, J-L, *Why there are Complementary Learning Systeme in the Hippocampus and Neocortex*, 1994

Meunier J-G, *Trois types de représentations.* Publication du LANCI. UQAM. 2002. www.unites.uqam.ca/lanci/

Murphy, *The big book of concept*, Massachusetts Institute of Technology. 2002

Piccinini: Voir Hampton [1993] pour les prototypes, Nosofsky [1988] pour les exemples, et Gopnik and Meltzoff [1997] pour les théories

Pierre Poirier et Guillaume Beaulac, *Le véritable retour des définitions*, Université du Québec à Montréal. 2010.

Popper, *La société ouverte et ses ennemis,*

Prinz, W., *Emerging selves: Representational foundations of subjectivit,* www.elsevier.com/locate/concog, 2003

Puzenat, Didier, *Parralléllisme et modularité des systèmes connexionnistes*, École normale supérieure de Lyon. 1997.

Quine, W. V. O., *Deux dogmes de l'empirisme*, in «Du point de vue logique. Neuf essais logico-philosophiques», Vrin, 2003

Rosch E., and Mervis, C. *Family resemblances: studies in the internal structure of categories*, Cognitive Psychology 7, 1975

Rosch, E., *Wittgenstein and categorization research in cognitive psychology*, in M. Chapman & R. Dixon (Eds.), *Meaning and the growth of understanding. Wittgenstein's significance for developmental*

Rowland, Mark, *The new science of the mind,* The MIT Press, 2010

Russell, B., *Human Knowkledge,* (1948)

Schmid A-F, *Concept, in Dictionnaire de Philosophie et des sciences.* Ed, PUF. Paris. 2003.

Shea Nicolas, *Content and Its Vehicles in Connectionist Systems, in. Mind & Language*, Vol. 22 No. 2007, pp. 246–269.

Simons et Chabris

Smolensky, *Le traitement approprié du connexionnisme.* in *Philosophie de l'esprit, problème et perspectives.* Texte réunis par Poirier et Fisette. Vrin. 2003.

Sperber, D. *Explaining Culture: A Naturalistic Approach.* Cambridge: Blackwell. 1996

Steel, *The symbol grounding problem has been solved*, so what's next?, 2008

Touzet C., *Introduction au modèle connexionniste.* 1992. dans google.

Varela, F.(1996) *Invitation aux sciences cognitives.* Paris. Édition du Seuil. 1ière éd(1989).

Vernant, J-P, *Mythe et pensée chez les Grecs. Études de psychologie historique*, Paris, François Maspero, 1965 ; rééd. Paris, La Découverte, 2007

Vernant, J-P, *La fabrique de soi,* Film de Emmanuel Laborie, 2011

Wilson, Robert, Collective memory, group minds, and the extended mind thesis, Cogn Process (2005)

Wittgenstein, *Investigation philosophiques*,

http://www.radio-canada.ca/nouvelles/Science-Sante/2006/08/24/001-pluton-planete.shtml