La représentation en philosophie de l'esprit et en sciences cognitives : la contribution de la phénoménologie et des neurosciences selon H. Dreyfus[1]

Sylvain Pronovost
Institute of Cognitive Science, Carleton University

[1] 1 Carleton University Cognitive Science Technical Report 2006-05. http://www.carleton.ca/iis/TechReports

I . Introduction

Le concept de représentation en philosophie de l'esprit est directement lié au problème de l'intentionnalité et de ses multiples interprétations et explications. Selon les différentes prétentions à la naturalisation des problèmes en philosophie de l'esprit, le néophyte est confronté à une quantité imposante de théories sur la nature de l'intentionnalité et de la conscience, qui sont par ailleurs incompatibles ou ne limitent qu'à de faibles espoirs les possibilités de concilier ces efforts[1]. Quel est le rôle de la représentation dans ces différentes explications? En général, le concept de représentation est un atout incontournable pour les explications sur la nature des relations sémantiques entre l'agent cognitif et le monde auquel il réfère.

Des penseurs comme W.V.O Quine, puis le «mouvement éliminativiste», refusant d'avoir recours à un dualisme conceptuel dans l'explication des propriétés des états mentaux selon les différences discriminées à leur égard face aux états physiques, suggèrent depuis un certain temps déjà l'abandon des concepts d'intentionnalité, de croyances, de désirs, et cetera. La plupart de leurs efforts se butent par contre eux aussi à des problèmes encore insolubles, comme le problème des niveaux de descriptions et d'explications (du comportement aux propriétés physiques du système nerveux), mais aussi dans la concrétisation de ces théories devenues modèles au sein des sciences cognitives, dont le projet de l'intelligence artificielle en est le témoin le plus évident.

Face aux échecs évidents du modèle fonctionnaliste classique que fut le computationnalisme en intelligence artificielle, le modèle connexionniste fut l'effort suivant le plus considérable en sciences cognitives pour fournir une base naturalisée forte à l'explication de la nature de la conscience et de l'intentionnalité. Hubert L. Dreyfus, qui fut longtemps et demeure le plus grand opposant à la possibilité de créer une intelligence artificielle, aborde le thème de l'intelligence dans un article intitulé *Merleau-Ponty's Critique of Mental Representation : The Relevance of Phenomenology to Scientific Explanation*, par une explication de la nature de l'apprentissage et du développement des

habiletés de l'homme **n'ayant pas recours à la représentation**. Selon son analyse, qui constitue une filiation originale entre la phénoménologie de M. Merleau-Ponty et les récents modèles des neurosciences, notamment la théorie de l'attraction de W. J. Freeman, il est possible et même souhaitable de conceptualiser le comportement intelligent sans avoir recours au concept de représentation, concept qui selon Dreyfus ne permet pas d'apprécier la réalité du développement cognitif de l'homme intimement lié à son contexte, à ses pratiques et à son expérience du monde.

Nous verrons en un premier temps ce que Dreyfus reproche au modèle fonctionnaliste classique au sein des sciences cognitives et de l'entreprise de l'intelligence artificielle computationnelle, puis les avantages et les limites du modèle connexionniste dans sa prétention à la reproduction de comportements intelligents. Dans un deuxième temps, nous prendrons en considération la conception «neurophénoménologique» de Dreyfus suite à la filiation qu'il crée entre la phénoménologie de Merleau-Ponty et le modèle connexionniste de Freeman en ce qui concerne l'apprentissage et le développement des comportements intelligents en général chez l'homme.

II . Du cognitivisme au retour à la phénoménologie

Dreyfus[2] conçoit l'intelligence artificielle comme un projet de transposition des prétentions rationalistes de penseurs tels Descartes, Leibniz, Kant et Husserl, en un programme de recherche empirique rigoureux. Le rationalisme et l'intelligence artificielle ont trois postulats essentiels en commun sur l'esprit : celui-ci est fondamentalement rationnel, il est constitué de représentations et est soumis à des règles. Pour les mêmes raisons que des penseurs tels Wittgenstein, Heidegger et Merleau-Ponty ont exposées comme critiques des limites du rationalisme, le projet de l'intelligence artificielle et celui du cognitivisme en général devraient être abandonnés, selon Dreyfus. Voyons d'abord en quoi consiste le modèle du cognitivisme classique avant d'apprécier la critique des limites de ce modèle comme l'entend Dreyfus.

Dreyfus s'oppose au modèle computationnel (de Turing et Von Neumann), dont la thèse sous-jacente est une théorie physicaliste et fonctionnaliste de l'esprit humain. Cette thèse fonctionnaliste repose à son tour sur ce que l'on nomme l'hypothèse symbolique de la GOFAI (*good old fashioned AI* - terme de Haugeland): le processus sous-jacent au comportement intelligent conçu comme étant de nature symbolique. Les trois présupposés traditionnels sur l'intelligence comprise comme architecture computationnelle («Von Neumannienne») selon la GOFAI étaient donc que celle-ci est constituée d'opérations sur des représentations symboliques abstraites, que ces computations sont gouvernées par un «programme» composé d'une liste de règles explicites de transformations opérant sur ces représentations symboliques, et que ces computations sont sérielles et sont exécutées par un «CPU» sur des informations conservées dans la mémoire permanente de «l'esprit ordinateur».

La GOFAI repose ainsi sur une théorie **représentationnelle** de l'esprit, où l'esprit est une entité qui exécute des calculs sur des représentations mentales (symboles) qui réfèrent au monde «externe». Cette conception de l'esprit en tant que dispositif de traitement sériel d'information symbolique, soumis à des règles, s'appuyait sur un fonctionnalisme de type fodorien, sur un «langage de la pensée» (LOT, language of thought de Fodor). De nombreux problèmes se présentèrent en raison de cette modélisation particulière de l'entreprise de l'intelligence artificielle et suscitèrent un grand nombre de critiques. Voici une brève liste des problèmes les plus significatifs :

- la GOFAI n'est que pure manipulation de symboles, donc n'opère qu'au niveau syntaxique. Ainsi, le programme ne «comprend» absolument rien (argument de H. Simon) puisqu'il ne peut renvoyer à des significations (sur ce point, voir aussi l'argument de J. Searle: le «Chinese room argument»).
- un micro-monde (un modèle utilisé en GOFAI pour résoudre des problèmes en utilisant un domaine restreint d'objets sur lesquels un programme peut opérer) n'est absolument pas représentatif de la relation cognitive établie entre l'agent et son monde. En fait, un micro-monde n'est pas une «subdivision mondaine» (référence à la conception phénoméno-ontologique de Heidegger), il ne réfère

qu'à un ensemble conceptuel fini, tel qu'utilisé en logique: un domaine, un univers, mais pas un **monde**... Un monde est un corps organisé d'objets, de buts, d'habiletés et de pratiques selon lesquels les activités humaines prennent sens, acquièrent des significations. Chaque«subdivision mondaine» présuppose déjà l'ensemble des pratiques et des objets du contexte général qui l'englobe et ne prend sens qu'en relation avec celui-ci...

- les descriptions conçues à partie de la simulation cognitive, soient des descriptions symboliques formelles, ne peuvent rendre compte de la complexité de la perception, par l'impossibilité de généraliser ces descriptions pour l'appréhension de cas particuliers. La *Gestalt theory of perception* (psychologie cognitive, cf. E. Goldmeier, 1972) suggère une explication neurophysiologique beaucoup plus adéquate à partir du concept de «Prägnanz», soit d'un phénomène physique de «résonance» des données de la perception sur des régions du système nerveux.
- la prédétermination des règles d'apprentissage d'un programme en vertu d'une sélection externe au programme (par le programmeur) des critères et aspects pertinents à l'assimilation de données fait en sorte que le programme n'apprend donc rien, puisque son appréhension est déterminée à l'avance...
- les descriptions des règles (critères) de définition de catégories dépendent du contexte et renvoient toujours à d'autres aspects, objets ou événements pour leur compréhension.

Le problème de la représentation du savoir fut le motif principal des tentatives de réajustement en intelligence artificielle pour l'appréhension du concept de représentation dans leurs futurs modèles. La psychologie cognitive (cf. E. Rosch, 1977) présentait déjà assez d'évidences empiriques démontrant que l'homme n'est pas conscient d'une classification d'objets en tant qu'instances de règles abstraites (devant convenir à un certain nombre de critères nécessaires et suffisants) mais qu'il regroupe plutôt les objets en fonction d'une similitude plus ou moins grande avec un paradigme imaginé, une image concrète d'un membre typique d'une catégorie. Le «knowledge-representation problem» (KRP) nécessitait donc d'une simulation cognitive qu'elle puisse appréhender tout fait pertinent dans des contextes différents...

Dreyfus reprocha à l'intelligence artificielle de ne pas tenir compte de l'aspect des attentes (*«expectations»*) de l'agent cognitif, comme l'entendait Husserl avec son analyse phénoménologique. L'intelligence, selon Husserl était déterminée par des contextes, constamment dirigée par des buts, comme une recherche de faits anticipés. Son «noème» est une représentation mentale située à l'intérieur d'un horizon d'attentes (*«predelineations»*) structurant les données sensorielles. Le noème est donc une description symbolique des aspects invariables de l'objet, à laquelle est ajouté un horizon de propriétés possibles mais non nécessaires de ce type d'objet. Mais une telle perspective phénoménologique posait un problème: les critères de sélection des objets pour l'établissement des stéréotypes appellent invariablement au contexte constitué par l'ensemble du savoir général de l'agent cognitif; le dénombrement des critères renvoie toujours à d'autres et constitue ainsi une tâche infinie... Husserl l'admit lui-même, son entreprise de définition des noèmes était vouée à l'échec. Heidegger fournit une explication beaucoup plus adéquate de la relation entre les horizons interne et externe de significativité, que Dreyfus reprend comme point de départ dans son argumentation : l'horizon des pratiques et de la culture générale de l'homme est la **condition** de possibilité pour déterminer les faits et aspects pertinents, et constitue ainsi un élément préalable à la structuration de l'horizon interne (la conscience structurante selon des attentes).

Ainsi, ce qui compte comme fait dépend toujours du contexte, mais ce contexte est déterminé par la **focalisation** actuelle de l'attention de l'agent cognitif ainsi que par ses **objectifs** (car l'homme est toujours «**déjà en situation**», contrairement au programme computationnel qui n'est dans aucune situation). Il est impossible de donner une description formelle exhaustive de ces conditions préalables à l'élaboration d'une description d'un contexte. Dreyfus souligna ainsi ce qu'il nomma le CKP, *«commonsense knowledge problem»* : un programme ne peut «comprendre» simplement par sa constitution, à partir d'une matrice très complexe de descriptions symboliques prototypiques, de schèmes «contextuels». Il ne possède pas de **savoir-faire** ni de connaissances du **sens commun** acquises par **l'expérience**. L'impossibilité de construire un modèle satisfaisant pour la compréhension du sens commun est causée par l'impossibilité de rendre compte des notions de:

- la pertinence des faits dont dispose un agent cognitif;
- la dépendance de la signification des faits selon le contexte;
- l'interaction de l'agent cognitif avec sa situation (contexte) donnant sens et pertinence aux faits et objets de son environnement.

Le problème de la pertinence («relevance») a donc source dans le problème du savoir du sens commun (CKP), lié au savoir-faire de l'homme; l'aspect fondamental de ce problème est sa relation à un contexte. Les données du contexte n'ont en effet elles-mêmes rien à offrir en terme de pertinence si elles sont prises en isolation du contexte. Ainsi, ce n'est qu'en fonction de l'expérience de l'homme dans un contexte de données que celles-ci peuvent être prises en considération (pertinence) et acquérir une valeur (signification holistique). Il en résulte une incompatibilité de la théorie représentationnelle de l'esprit avec la connaissance du sens commun (CKP) : le modèle computationnel, reposant sur la thèse que l'esprit est une structure informationnelle de représentations symboliques atomiques et discrètes sur laquelle opère un ensemble de règles (soit un modèle logico-fonctionnel de la signification), est absolument incompatible avec les arguments avancés par Dreyfus. Le modèle computationnel ne tient pas compte de la pertinence, du contexte et donc du caractère **holistique** de la signification dans la compréhension du sens commun.

Le savoir-faire (know-how), unique à l'homme, est cette faculté de généraliser et de déterminer la pertinence des données qui lui sont accessibles dans un monde constamment confronté, i.e. dans lequel l'homme est «toujours-déjà-en-situation» (cf. Heidegger dans *Sein und Zeit*). La GOFAI ne peut rendre compte du savoir-faire de l'homme, qui dépend de l'expérience directe de l'homme avec son monde, par l'intermédiaire essentiel du corps. Seul le savoir-faire permet de juger de la pertinence des faits portés à la conscience de l'homme, et ne dépend pas de la manipulation de données, aussi nombreuses peuvent-elles être.

Sur un tout autre aspect problématique du modèle computationnel, Dreyfus s'intéressa au traitement sériel des symboles et à la réalité biologique de la cognition. Il apparaît impossible que l'information modélisée dans le modèle computationnel de l'intelligence soit représentative du traitement réel de l'information par le cerveau. Le traitement sériel du modèle computationnel ne rend pas compte de la disposition **combinatoire** de l'information composant les données sensorielles, vraisemblablement la disposition réelle du traitement de l'information par le cerveau de l'homme. Le PDP (parallel distributed processing) ou connexionnisme, pouvait rendre compte avec plus d'efficacité du traitement **parallèle** de l'information massive à laquelle l'homme est confronté. Ses avantages sont nombreux :

- il est parallèle et non sériel;
- il utilise une formalisation mathématique de l'information sous forme d'assignations de valeurs analogiques (continues), donc non computationnel (composé de valeurs discrètes, dans un formalisme logique);
- il est vraisemblablement biologiquement adéquat (système complexe de processus parallèles du traitement de stimuli cherchant un équilibre neuronal constant).

Pour en revenir au problème de la représentation du savoir, il faut souligner que Dreyfus soutient que ni la GOFAI, ni le PDP ne peuvent permettre de résoudre le problème du savoir du sens commun (CKP). Selon lui, que le modèle de l'intelligence soit computationnel ou connexionniste, cela n'élimine en rien le CKP, conçu sous ses trois dimensions:

- comment le savoir «de tous les jours» est-il organisé pour que l'on puisse en tirer des inférences?
- comment les habiletés et savoirs peuvent-ils être représentés comme «savoir-que», alors qu'ils devraient plutôt être conçus comme «savoir-comment» (know-how)?
- puis, comment le savoir pertinent peut-il mis à profit dans des situations particulières?

Dreyfus en appelle d'une conception que je nommerai un **holisme ontologico-sémantique** à partir d'un point de vue **phénoménologique**, pour expliquer le comportement intelligent de l'homme. La capacité cognitive supérieure de l'homme face à toute simulation cognitive ne provient pas d'un degré de complexité du système représentationnel qui n'a pas encore été atteint; c'est le rapport phénoménologique de l'homme, toujours «en situation», avec le monde, qui lui permet d'entretenir un savoir-faire, en termes de compétences et de connaissances. La pertinence et la signification ne sont envisageables qu'en fonction d'une «contextualisation» constante de l'homme, et ne sont donc pas accessibles à la simulation cognitive. En ce qui concerne l'ordinateur : 1- en l 'absence de contextes dans lesquels il pourrait faire l'expérience et la sélection de faits, événements et objets pertinents, il traite la valeur des faits qui lui sont soumis comme étant toujours la même. 2- Tous les faits disponibles à la simulation cognitive étant présélectionnés, ils sont donc toujours pertinents et doivent être testés un par un. 3- Le problème est qu'il existe un nombre aleph de faits: nous avons ainsi le choix de donner à la simulation cognitive ou bien TOUS les faits sous forme de représentations, ou bien en exclure un certain (immense!) nombre. 4- De cette présélection (exclusion) de faits, un grand nombre doit être invariablement pertinent selon les circonstances...

L'intelligence artificielle et le cognitivisme en général, en faisant appel au concept de représentation dans l'explication des comportements intelligents de l'homme, sont donc victimes de la circularité de l'entreprise de définition des concepts de pertinence et de contexte :

1- Que le modèle soit computationnel ou connexionniste, tenter de pourvoir ce modèle de critères de sélections selon la pertinence est un piège de régression circulaire, puisque les critères de pertinence dépendent du contexte,

2- et tenter de pourvoir ces modèles de critères de définition des contextes dépend des notions jugées pertinentes préalablement à l'appréhension de ces contextes...

Ainsi, c'est parce que la **simulation** cognitive n'entretient pas de mise en situation préalable et constante qu'elle ne pourra jamais être une **reproduction**

cognitive (cf. encore J. Searle et son «Chinese room argument»)...

En guise de conclusion de cette première partie, nous pouvons dégager quelques points de l'argumentation de Dreyfus quant aux problèmes inhérents au projet strictement fonctionnaliste du computationnalisme au sein des sciences cognitives et plus particulièrement dans l'entreprise de l'intelligence artificielle. Suivant l'impasse des micro-mondes en vertu de laquelle les descriptions symboliques dans un modèle fonctionnaliste de l'esprit ne peuvent rendre compte des faits de l'existence mondaine disponibles à la cognition humaine, les cognitivistes ont dû prendre en considération les notions de contexte et de pertinence. Malheureusement, suivant le même **présupposé métaphysique** (cf. Heidegger) qui a poussé le rationalisme à dichotomiser le monde externe et la vie psychique (où tout n'est que représentations par abstractions selon des règles), les cognitivistes ont effectué une **seconde erreur** (fort prévisible lorsque l'on prend le recul du point de vue phénoménologique): traiter le contexte lui-même et ses caractéristiques (pertinence et signification) comme un simple **objet** circonscrit par une description structurée...

Dreyfus, se disant antiformaliste et antimécaniste, tient la thèse suivante: en référence à Wittgenstein et à sa conception des jeux de langage, il soutient que l'analyse du comportement humain en termes de règles contient toujours une condition **caeteris paribus**. Ainsi, les règles s'appliquent «toute autre chose étant équivalente», soit que dans une situation spécifique, toute «chose étant équivalente» ne peut être décrite sans régression contextuelle infinie... Dreyfus retient deux hypothèses contre le modèle computationnel. La première fait référence à la phénoménologie de Merleau-Ponty et à la philosophie du langage ordinaire de Wittgenstein : la connaissance des intérêts et pratiques de l'homme **ne peut être représentée**, i.e. le savoir en général est acquis par l'expérience constante de l'agent cognitif dans un contexte, sans possibilité de le décrire en isolation de ce contexte, tout comme le savoir-faire (apprendre à nager, par exemple) est acquis par l'expérience de la relation corporelle de l'agent au monde. La seconde hypothèse est que les représentations ne sont pas formelles et abstraites, mais plutôt picturales et concrètes (suivant la psychologie gestaltiste), issues de l'expérience et non soumises à des règles strictes et à la condition caeteris paribus de celles-ci.

L'échec de la GOFAI et du cognitivisme en général serait ainsi imputable à la négligence de l'aspect phénoménologique de la relation de l'agent cognitif à son monde. Le point principal peut être résumé ainsi : puisque l'intelligence doit être **située**, elle **ne peut être séparée du reste de la vie humaine**. Mais cette erreur ne revient pas à la seule entreprise de l'intelligence artificielle: la tradition philosophique, depuis Platon jusqu'au projet rationaliste des Descartes, Leibniz, Kant et Husserl, s'est toujours évertuée à distinguer l'esprit du corps et le théorique du pratique. Le savoir-faire et la connaissance commune des agents cognitifs que nous sommes sont directement liés à nos habiletés motrices et sensorielles par exemple, dans le développement de notre faculté de discrimination et d'adaptation aux objets; à nos désirs et besoins dans la structuration de notre situation sociale; puis encore à notre contexte culturel influençant l'interprétation individuelle impliquée dans le savoir-faire que nous manifestons. *«Great artists have always sensed the truth, stubbornly denied by both philosophers and technologists, that the basis of human intelligence cannot be isolated and explicitly understood...»*[3]

III . Neurophénoménologie et intelligence : contre le concept de représentation dans l'explication de l'apprentissage

La position de Dreyfus face au connexionisme a changé depuis l'évolution des modèles neurobiologiques, dont les résultats sont de plus en plus fructueux. En fait, Dreyfus refuse toujours la possibilité d'une intelligence artificielle en vertu de sa perspective phénoménologique de la cognition, mais les récents modèles connnexionnistes possèdent un avantage que l'on ne saurait négliger : ils proposent une approche empirique de l'explication du comportement intelligent et malgré leurs lacunes, ils remplissent parfaitement bien le modeste rôle qu'il leur revient, soit d'être des modèles plus ou moins précis du fonctionnement et de la nature neurobiologique de facultés telles la perception, l'apprentissage, la mémoire, la classification, et cetera.

Dreyfus propose, dans son texte *Merleau-Ponty's Critique of Mental Representation : The Relevance of Phenomenology to Scientific Explanation*

(1998), de mettre en relation les acquis de la phénoménologie de Merleau-Ponty sur la perception, l'apprentissage, l'intelligence et leur dépendance au corps, avec l'explication neurobiologique des mêmes facultés selon W. Freeman. Ainsi, en liant ces deux approches très différentes du même champ d'investigation, Dreyfus espère démontrer une **isomorphie** tout à fait particulière entre ces différents types d'explication en jeu. De plus, comme le souligne le sous-titre et thème principal de l'article, *Intelligence without Representation*, Dreyfus prétend donner une explication de l'apprentissage de connaissances théoriques et de l'acquisition d'habiletés requérant des compétences cognitives propres à l'homme **exempte du concept de représentation**, du moins au sens fort du terme. Voyons maintenant en détails ce que Dreyfus entend par l'apprentissage et l'intelligence exempts de représentations.

Deux concepts importants issus de la P*hénoménologie de la perception* de Merleau-Ponty suscitent une attention particulière dans l'explication de l'apprentissage selon Dreyfus : les notions «**d'arc intentionnel**» et de «**prise maximale**»[4], soit la connexion liant intimement l'agent à son monde (l'agent développe des habiletés «emmagasinées» dans le corps, non pas sous forme de représentations, mais plutôt sous forme de dispositions du corps à répondre selon la sollicitation de diverses situations), et la tendance du corps à répondre à ces sollicitations de façon à amener la situation vécue le plus près possible du sens du **gestalt**[5] optimal de l'agent.

Selon Dreyfus, ces deux phénomènes ne requièrent pas le recours au concept de représentation mentale et sont isomorphiques aux modèles de réseaux neuronaux utilisés pour rendre compte de la perception et du développement de compétences. Merleau-Ponty affirme que la relation de l'agent à son monde lui permet d'acquérir des habiletés, et ces habiletés déterminent en retour comment les situations se présentent à l'agent de façon à requérir de lui des réponses. Suivant cette idée, Dreyfus propose d'étayer l'établissement de l'arc intentionnel en décomposant le processus d'acquisition de connaissances théoriques et d'habiletés physiques à l'aide d'exemples concernant les échecs (le jeu de stratégie) et la conduite automobile[6]. Il faut

souligner que ces processus sont décrits de façon purement indicative et Dreyfus ne prétend pas discriminer catégoriquement les différentes étapes du développement d'habiletés et de l'apprentissage de connaissances théoriques. Les points qui suivent constituent une brève description du processus d'acquisition d'habiletés d'un adulte par instruction (pour éviter la simplicité d'une explication de type essais et erreurs au cours du développement des enfants), du point de vue phénoménologique de Dreyfus :

-1- **novice** : l'agent est confronté à une nouvelle activité et a recours à un ensemble de règles de base indicatives hors de tout contexte d'utilisation. Ainsi, le débutant doit se soumettre à des instructions déjà définies et ce, dans l'absence de contexte (une expérience très restreinte de l'activité en cause).

-2- **débutant avancé** : la mise en situation des acquis par les règles permet à l'agent de ce stade du développement de l'habileté de discriminer de nouveaux aspects, cette fois-ci contextuels, de l'activité. Cette discrimination d'aspects pertinents mais contextuels supplée aux situations dont les descriptions sont très intimement liées à des contextes précis et échappent à l'explicitation.

-3- **compétence** : l'individu à ce stade de développement d'une habileté reconnaît désormais bien plus d'éléments pertinents et plus de nuances dans les diverses possibilités qui s'offrent à lui dans l'exercice de cette habileté. Il adopte désormais des stratégies précises parmi un vaste nombre d'approches d'un même problème, se basant sur son expérience croissante des contextes où il requiert cette habileté. La compétence permet d'exercer une habileté de manière à sélectionner aisément les aspects pertinents à la résolution d'actions, mais ces choix sont encore restrictifs, puisque l'agent n'a pas encore la capacité de reconnaître tous les aspects pertinents. À ce stade se développe aussi le sens de l'erreur et de la responsabilité dans l'exercice de l'habileté, et Dreyfus met l'emphase sur l'engagement émotionnel qui se développe parallèlement à une compétence, selon les succès et les échecs conséquents.

-4- **talent** (*«proficiency»*) : à ce stade, la performance de l'agent est de plus en plus intuitive. Il discrimine désormais une grande variété de situations par l'expérience acquise et peut associer la réponse optimale non plus par le calcul

des alternatives qui lui sont offertes, mais par la discrimination du choix le plus adapté à la situation actuelle. Les réponses intuitives ne demandent pas l'exercice d'un jugement détaché, source de doute et de prudence. Malgré sa spontanéité croissante, la personne talentueuse ne sait pas toujours quelle réponse associer à une situation bien qu'elle soit en mesure de discriminer ses objectifs précis. Elle voit plus de réponses possibles que de situations possibles, et doit parfois se rabattre sur les règles pour adopter une décision optimale.

-5- **expertise** : Alors que la personne talentueuse voit quoi faire et décide comment réaliser son objectif, l'expert, par l'expérience considérable de son habileté ou de son savoir, voit ce qu'il doit faire et fait ce qui est requis dans une multitude de situations. L'expert discrimine les situations les plus subtiles et procède par réflexe, selon le vaste répertoire de configurations et de stratégies acquises par l'expérience (un champion international aux échecs peut distinguer jusqu'à environ 50 000 types de configurations!). Ainsi, l'expert procède de façon si intuitive qu'il associe spontanément la réponse appropriée à la situation actuelle, il ne s'appuie plus sur la conscience et le calcul des possibles. L'immédiateté de ses réponses suggère une capacité cognitive impressionnante de subdivision et de structuration de classes de situations, où l'analyse n'a plus sa place et l'agent peut s'exécuter avec rapidité sans subir de dégradation de performance.

Dreyfus affirme que l'histoire précédente sur le développement d'habiletés et l'apprentissage de connaissances suggère une conception non représentationnelle de l'apprentissage. Merleau-Ponty soutient que ce n'est que par l'expérience du corps seule que se constituent des associations de situations, d'événements, d'interactions avec des objets : des impressions ne renvoient jamais à d'autres par elles-mêmes, et le recours à une conception de la mémoire comme matrice de représentations des situations vécues antérieurement est obsolète. Les situations se présentent plus précisément et avec plus de détails à l'agent expérimenté, et sollicitent en retour des réponses plus rafinées de sa part. Ainsi, l'apprentissage apparaît dans la façon dont le monde se présente à celui qui a appris.

Cette rétroactivité dans la relation dynamique que l'agent entretient avec son environnement est justement ce qui définit la notion d'**arc intentionnel** chez Merleau-Ponty. Ainsi, il n'est pas question d'une relation passive entre la perception de stimuli et le renvoi de réponses par l'intermédiaire de l'agent : **l'agent est toujours en situation de réponse sollicitée**. Une réponse motivée par un apprentissage de la part d'un agent dépend toujours de son expérience dans des situations concrètes semblables. Cette expérience est projetée rétroactivement dans la perception du monde. Nul besoin de cet intermédiaire indésirable qu'est la représentation mentale donc, dans l'explication de l'apprentissage : Dreyfus affirme ironiquement que la meilleure représentation du monde est le monde lui-même...

Préoccupons-nous maintenant de la filiation qu'effectue Dreyfus entre la phénoménologie non représentationnelle de Merleau-Ponty et la modélisation connexionniste des mécanismes neurobiologiques sous-jacents à la conscience. De récents modèles neuronaux (*feed forward simulated neural networks* et *recurrent neural networks*, que je simplifierai à l'abréviation NN, pour *neural networks*), démontrent d'étonnantes caractéristiques du système neuronal au sein d'un système complexe et dynamique qui ne néglige pas l'interaction entre l'agent virtuel et son environnement comme l'ont malheureusement fait des modèles antérieurs, d'inspiration empiriste ou rationaliste. Selon ces récents modèles connexionnistes, le NN, bien plus qu'un simple intermédiaire fonctionnel associant des réponses à des stimuli, possède des propriétés biologiques que l'on ne peut négliger dans l'explication du comportement intelligent. Entre autres, c'est un système rétroactif qui ajuste l'assimilation des stimuli perçus en fonction d'un certain apprentissage contextuel suggérant des réponses appropriées. Au concept de mémoire se substitue dès lors celui de «**configuration**» de l'information, emmagasinée dans les multiples connexions entre les neurones et constituant l'acquis du système face à son expérience perceptive et motrice. C'est par l'ensemble formé par les stimuli et l'état initial du NN que sont déterminées les réponses, et cet état initial du NN est donc une «configuration» de l'expérience du système (du corps chez l'homme) où les concepts de

mémoire et de représentation semblent curieusement obsolètes. Ces considérations suggèrent aussi une impressionnante isomorphie avec la notion d'arc intentionnel de Merleau-Ponty, sur laquelle nous reviendrons ultérieurement.

Malgré les quelques avantages énumérés précédemment, certains problèmes limitent un tel modèle. Dreyfus souligne entre autres que :

-1- Le pairage des stimuli et des réponses pour l'apprentissage des NN est effectué antérieurement par les programmeurs, mais des efforts pour rendre les NN autonomes à l'aide de fonctions rétroactives donnent déjà des résultats positifs.

-2- Pour être reconnue comme intelligence, la capacité cognitive d'un NN, en vertu de son caractère artificiel, doit être similaire au niveau comportemental aux compétences des êtres humains. Le problème est qu'il existe un nombre indéfini de perspectives sur les similitudes et les différences du nombre infini de stimuli.

-3- Ce qui amène Dreyfus à ce qu'il nomme le problème de la **généralisation** : quels sont les critères de sélection des associations faites par une communauté d'agents intelligents tels que nous le sommes? Face à ce problème d'interprétation qui est déjà le propre des individus d'une même espèce comme la nôtre, comment espérer qu'une intelligence reproduite artificiellement pourrait développer les mêmes compétences que les êtres humains?

Merleau-Ponty donne des indications sur la faculté d'association et de généralisation qui peuvent mettre à l'épreuve les problèmes précédents, et ces indications sont de surcroît confirmées par le travail de W. Freeman en neurobiologie. Selon Merleau-Ponty, le problème de la similarité n'a pas source dans la comparaison de représentations. Il s'agit plutôt d'une discrimination par prototypes des stimuli (confirmée par la psychologie cognitive gestaltiste comme nous l'avons vu précédemment, dans la première partie de l'exposé) et non pas une association de représentations, soit une appréhension par le cerveau à l'aide de ses fonctions perceptives de la distance plus ou moins grande entre les stimuli pris dans leur contexte et le gestalt (configuration prototypique) formé par l'expérience du corps.

Selon Dreyfus, l est inconcevable, comme pour Merleau-Ponty, de négliger l'incarnation essentielle du comportement intelligent (l'intelligence dans le corps) si nous voulons expliquer le phénomène de généralisation. Dreyfus distingue trois aspects de la configuration du corps pour tenter d'expliquer la faculté de généralisation propre aux organismes complexes que nous sommes :

-1- La généralisation dans l'association de stimuli contraignant les réponses possibles est d'abord restreinte en vertu de **l'architecture cérébrale** que nous possédons. En effet, les contraintes imposées par les «constantes» perceptives du système nerveux ordonnent un premier niveau d'association fondamental des stimuli, qui est donc dépendant de la structure du cerveau.

-2- Ensuite, **l'ordre** et la **fréquence** dans la présentation des stimuli contraignent aussi la généralisation, et dépendent de **l'interaction de la structure du corps avec la structure du monde**.

-3- Enfin, l'aspect **téléologique** de la **satisfaction recherchée par le corps** contraint aussi la généralisation dans l'association de stimuli et de réponses. Ainsi, le pairage de ces derniers se fait en fonction d'une certaine mesure de ce qui compte en tant que succès.

Dreyfus suggère que ces trois fonctions structurantes du corps pourraient bien être suffisantes pour expliquer la faculté de généralisation des êtres humains, et conséquemment la faculté d'apprentissage et du développement d'habiletés. Le problème fondamental de l'intelligence artificielle se précise donc davantage que dans la première partie de l'exposé : sans corps, un NN artificiel est essentiellement désavantagé dans la prétention qu'on lui attribue à l'apprentissage, puisqu'il ne peut distinguer les généralités issues de la perception, de la motricité, de la séquentialité, de l'ordre des stimuli, ni même un quelconque aspect téléologique; il ne peut produire de réponses émotionnelles face à ses succès et échecs, et cetera. Dans cette perspective, il apparaît évident que le corps soit une condition **essentielle** au comportement intelligent...

Un autre problème concernant la représentation se pose lorsque l'on considère ce qui a été dit précédemment sur l'aspect téléologique du l'intelligence incarnée. En effet, comment peut-on affirmer que l'on n'a pas

besoin de recourir au concept de représentation pour apprécier l'aspect téléologique de l'intelligence incarnée, si celle-ci doit avoir à sa disposition une certaine mesure du succès ou de l'échec face à ses objectifs? Pour Merleau-Ponty, une action peut satisfaire des conditions téléologiques sans que l'agent les ait à l'esprit comme objectif. Selon lui, l'activité orientée téléologiquement est **une relation intentionnelle plus fondamentale** que le type de relation intentionnelle faisant appel aux représentations. La relation du corps au monde est beaucoup plus contraignante que la conception des «représentations objectives» issue du rationalisme : Merleau-Ponty fait appel à la notion de **prise maximale** pour expliquer cette relation encore plus stricte et sous-jacente à sa conception de l'arc intentionnel constitué par la rétroaction entre le corps et le monde.

La prise maximale dénote cette tendance des organismes vivants à obtenir une appréhension optimale de la situation dans laquelle ils se retrouvent. Ainsi, notre corps cherche par exemple toujours à obtenir les conditions optimales de perception de son environnement, et tend à s'adapter par des actions appropriées à cet environnement de façon à en tirer une satisfaction optimale. À partir de sa conception de la prise maximale, Merleau-Ponty en vient à la conclusion qu'il n'est nullement nécessaire de se représenter nos objectifs : la réalisation d'une activité se fait à partir (1) d'une **déviation initiale du gestalt optimal**, constitué par la relation rétroactive du corps et du monde, puis (2) de la **tension sollicitant le corps au rétablissement d'un équilibre** entre le corps et ce gestalt soulageant cette «tension». Cette recherche du gestalt optimal n'a pas besoin d'être «représentée dans l'esprit», la tension sollicitant le corps à le recherche d'un équilibre face à une situation précise peut être perçue ou non, ce n'est pas nécessaire à sa réalisation.

En fait, le «sens» de l'optimal et de la déviation de ce gestalt-équilibre ne relève pas de la pensée consciente, il est inscrit dans le corps compris en tant que système dynamique. Dreyfus souligne que la pensée consciente n'est pas «causale», elle ne sollicite pas l'action, prise en isolation du corps. C'est l'**expérience** du corps visant un gestalt optimal qui est **causale**, soit l'interaction

rétroactive entre le **corps et le monde**. Un autre aspect fort intéressant de ce phénomène de prise maximale dans l'explication de l'apprentissage est que l'agent cherche non seulement le gestalt optimal dans de multiples situations (l'amenant à «apprendre» à associer des réponses à des stimuli), il **tend aussi à améliorer ce gestalt optimal lui-même**. Ce deuxième aspect, le raffinement de l'arc intentionnel avec l'expérience, rend très bien compte de ce que l'on entend par un développement optimal de comportements intelligents et d'habiletés. Toutefois, ce double aspect téléologique demeure non représentationnel, car il implique d'abord et avant tout que l'agent soit impliqué activement et s'exerce rigoureusement dans l'apprentissage d'une compétence quelconque s'inscrivant dans le corps, et non à partir d'une matrice de représentations mentales. Comme nous l'avons dit précédemment, cette activité peut être consciente ou non, mais il demeure que c'est à partir du phénomène de prise maximale, selon Merleau-Ponty, que s'établit la relation stricte de l'arc intentionnel entre le corps et le monde.

Que le corps réponde au monde ne fait pas de problème en soi dans une conception non représentationnelle de la relation corps-monde; c'est plutôt **l'orientation** de la relation causale qui semble problématique. Ainsi, une conception représentationnelle de la cognition soutient généralement que la relation causale s'établit dans une direction «esprit-vers-monde», alors que l'approche phénoménologique de Merleau-Ponty soutient l'opposé, soit que l'orientation de la relation causale de l'action, intelligente ou instinctive, est de type **«monde-vers-esprit»**. Il s'agit ici de distinguer entre un agent intentionnel pris isolément d'une part, et la sollicitation par un gestalt-équilibre optimal d'une réponse du corps aux stimuli de multiples situations. Nul besoin de s'en faire avec le pseudo-problème de l'action passive et non intentionnelle, où il ne serait pas possible de distinguer entre une action causée par l'agent et une autre causée par des facteurs externes. L'explication de Merleau-Ponty se situe dans une toute autre perspective : l'expérience du monde sollicite la réponse du corps et ce, dans la relation rétroactive (l'arc intentionnel) composée par ces deux éléments, le corps et le monde.

Pour résumer la position phénoménologique de Dreyfus, qui fait appel à celle de Merleau-Ponty, la réponse du corps est une action si et seulement si :
-1- l'agent contrôle son action de façon à pouvoir interrompre celle-ci;
-2- celle-ci est par contre causée par le gestalt optimal, soit la relation rétroactive entre corps et monde tendant vers la prise maximale. L'action est ainsi «sollicitée» par une mise en situation continuelle.

Bien sûr, Dreyfus ne contredit pas les conditions logiques d'une action dans le cas d'une mise en situation nouvelle pour un agent, soit que l'action est dans ce cas d'abord motivée par une réflexion sur l'expérience de façon plutôt objective et détachée. La **délibération** sur une action ne relève par contre que d'une analyse logique de l'expérience et néglige l'interaction fondamentale de la dynamique créée par l'agent et son contexte. Dreyfus souligne que la «délibération» sur une action n'entre en jeu que lorsque celle-ci ne tend incidemment pas à la prise maximale du gestalt-équilibre, et repose elle-même sur le fond d'une recherche optimale d'adaptation aux situations. Malgré toutes les explications précédentes, Dreyfus est conscient que l'aspect téléologique non représentationnel de l'action dans la perspective phénoménologique de Merleau-Ponty demeure un mystère et échappe encore à une explication précise. La fonction téléologique peut-elle amener un agent à rechercher un équilibre sans que celui-ci n'ait vraiment aucune idée de ses objectifs? Dreyfus donne l'exemple du jeu d'essais et d'erreurs où un joueur indique à l'aide d'indices à un autre joueur si ce dernier se rapproche ou s'éloigne de la solution recherchée. Dans cet exemple, le joueur qui cherche n'a pas connaissance de son objectif et s'y dirige pourtant, mais l'autre joueur doit savoir où le mener pour que l'activité soit fructueuse. Une explication scientifique de ce type de fonction téléologique sans représentation de l'objectif est-elle possible? Dreyfus se tourne du côté de la neurobiologie et du modèle connexionniste...

La théorie de l'attraction de W. Freeman pourrait être la solution à notre problème. Comme nous l'avons mentionné précédemment, au sens le plus faible du mot, la «représentation» de la situation optimale sans connaissance abstraite et symbolique de l'objectif ciblé est possible dans un modèle de psychologie

gestaltiste. La contribution de Freeman est de proposer qu'à partir du modèle Hebbien d'une théorie de l'apprentissage d'un NN par l'ajustement rétroactif des forces de connexions entre neurones basé sur l'expérience, il est possible d'expliquer le fondement d'un processus de niveau qualitatif supérieur. L'apprentissage serait ainsi structuré par le renforcement de l'expérience contraignant les stimuli, créant un bassin spécifique «d'**attraction**» des stimuli associant ceux-ci à des réponses semblables. Il faut souligner que ce modèle sous-entend une mise en situation directe du NN de l'homme avec des objets et des événements, créant une relation directe et stricte entre le système nerveux et son environnement. Ainsi, c'est par la répétition de la perception de stimuli semblables que se créent des «bassins attracteurs» provoquant des réponses semblables, par la contrainte de la configuration des associations possibles. Le NN passe ainsi du chaos de la multiplicité de stimuli à l'ordre des associations qui constitue «l'apprentissage».

Qu'en est-il du phénomène de prise maximale de Merleau-Ponty? Par l'expérience de succès et d'échecs répétés dans des situations semblables, le cerveau forme des connexions dont la configuration renvoie à l'assimilation des associations de stimuli et de réponses. Cette configuration du NN cherche un état d'activation minimale, pour se libérer de la tension de l'activation maximale suscitée par une situation nécessitant l'activation du corps d'une certaine façon. La configuration du NN, dans une situation analogue à des expériences précédentes mais jamais exactement la même, tend à renforcer ses associations entre stimuli et réponses semblables, par son activité spécifique au niveau du flux énergétique qui le parcourt. Les actions causées par l'interaction entre le NN et ses stimuli tendent dès lors à **ramener l'état du système nerveux à l'influx énergétique minimal**, soit par attraction vers le bassin (le niveau de la configuration) suscitant l'équilibre le moins coûteux en terme d'activation. Au niveau phénoménal, nous percevons cette tendance du corps à nous ramener à l'état d'activation minimale par la recherche du **gestalt-équilibre optimal** dans la sollicitation du corps par la situation. Ainsi, le système est toujours en état

d'activation plus ou moins distant de l'influx énergétique optimal, soit le bassin attracteur où l'énergie requise est minimale dans la configuration du NN.

Au niveau **neurobiologique**, le système oriente ainsi l'agent à l'activation énergétique minimale dans la configuration de son «expérience» assimilée par le renforcement du réseau neuronal. De la même façon, le système oriente aussi l'agent à l'atteinte du gestalt optimal d'une situation spécifique au niveau **macrophénoménal**, qui cause cet état d'excitation, de tension dans le corps. Ce mécanisme neurobiologique et cette interaction phénoménale corroborent ainsi une faculté d'adaptation de l'organisme, que nous appelons apprentissage et développement d'habiletés, **sans le recours aux représentations mentales**. L'agent ne se représente pas le point d'activation minimal et l'équilibre optimal dans une situation où il doit agir, pour atteindre ses fins. Il tend à cette fin sans jamais avoir à la connaître, selon une relation rétroactive avec son environnement, son monde. «L'attracteur» de Freeman et la prise maximale de Merleau-Ponty ne pourraient être qualifiés de représentations qu'au sens le plus faible de phénomène où l'expérience passée est prise en considération par le système nerveux de l'agent, dans l'atteinte de ses objectifs sollicitant ses actions. Dreyfus considère donc qu'il y a isomorphie entre la conception phénoménologique non représentationnelle de Merleau-Ponty et le modèle neurobiologique de Freeman, dont le considérable avantage est d'offrir la possibilité d'une réconciliation entre les descriptions et explications de niveau phénoménal et neurophysiologique concernant la nature et le fonctionnement de l'esprit.

IV . Conclusion

Le projet de l'intelligence artificielle, compris comme modélisation fonctionnelle et représentationnelle de l'esprit humain, sous le schème du computationnalisme, était incapable de rendre compte de l'esprit selon ses diverses capacités essentiellement liées à la contextualité du corps dans le monde. Comme le rationalisme, la GOFAI posa le sujet de la cognition hors de toute situation, concevant l'esprit de façon autonome et isolée représentant son

monde symboliquement par abstraction, duquel il est d'ailleurs en «distance» puisque celui-ci est conçu en tant qu'objet posé par le sujet des représentations. Qui plus est, cet esprit devait être soumis à la gouverne de règles logico-formelles, où tout est inférence et causalité nomologique. L'esprit y était donc pure rationalité, sa structure même étant indépendante des affects, événements, objets et situations du monde auquel il est pourtant continuellement confronté.

Les prétentions du rationalisme et de la GOFAI doivent être rejetées selon Dreyfus, car elles ont continuellement négligé le caractère essentiellement holistique de l'existence de la conscience de l'homme en rapport avec le monde duquel il fait partie. Le recours aux concepts de représentations symboliques et abstraites, de règles logico-formelles et d'indépendance du sujet rationnel est obsolète et même nuisible dans l'effort de compréhension de la nature et du fonctionnement de l'esprit. Dreyfus propose tout simplement d'éliminer des concepts dont aucun théorie à prétention heuristique ne peut se satisfaire depuis les débuts mêmes de la philosophie de l'esprit. Une perspective phénoménologique est nécessaire, même essentielle à l'explication de la nature de l'esprit, puisque les thèmes principaux eux-mêmes de la philosophie de l'esprit et des sciences cognitives ne peuvent être compris qu'en relation avec un contexte, une expérience du monde englobant le sujet de la cognition. La conscience est toujours ***conscience de*** et l'intentionnalité est ***toujours intention de***, où l'objet de cette conscience et de cette intentionnalité ne peut être isolé dans leur étude.

Par une analyse originale de deux niveaux de description et d'explication forts différents de l'esprit, l'un phénoménologique et l'autre neurobiologique, Dreyfus s'est évertué à satisfaire des conditions épistémologiques et méthodologiques essentielles de la philosophie de l'esprit et de sa contribution aux sciences cognitives. Cet effort est d'autant plus original qu'il constitue un type d'explication de la nature et du fonctionnement de l'esprit qui combine une prétention à la naturalisation d'une part avec l'explication neurobiologique, et une

prétention philosophique ontologico-phénoménologique d'autre part, qui échappe au physicalisme radical mais y est pourtant isomorphe.

[1] Je fais référence à D. Dennett et à son travail de division des trois «postures» intentionnelle, architecturale et physique, à partir duquel se pose déjà la question de la compatibilité des niveaux d'explications de la conscience et de l'intentionnalité.

[2] in What computers still can't do: a critique of artificial reason (1993)

[3] *ibid.*

[4] «Intentional arc and maximum grip» dans le texte de Dreyfus.

[5] Le gestalt (du mot forme, ou encore structure, en allemand) est compris ici en tant que «configuration» prototypique concrète qui serait en cause pour expliquer la perception des stimuli sensoriels.

[6] Pour les besoins de la dissertation, je n'exposerai pas les exemples accompagnant chaque stade du développement des habiletés prises en considération par Dreyfus.