

# **Le naturalisme en philosophie de l'esprit et la critique du fonctionnalisme**

Tom Dedeurwaerdere

Université catholique de Louvain (UCLouvain) and F.R.S.-FNRS, Belgium

## **Bibliographical reference**

Dedeurwaerdere, T., 2008, "Le naturalisme en philosophie de l'esprit et la critique du fonctionnalisme", in D. Vanderveken and D. Fisette. *Actions, rationalité et décision. Action, Rationality and Decision*, London: College Publications, 203–230.

## **Self-archived author copy**

This copy is for your personal, non-commercial use only.

For all other uses permission shall be obtained from the copyright owner.

**Copyright:** © College Publications- All rights reserved.

## 8

## Le naturalisme en philosophie de l'esprit et la critique du fonctionnalisme

TOM DEDEURWAERDERE (FNRS, UNIVERSITÉ DE LOUVAIN)

L'objet de cette contribution est une réflexion sur les conséquences pour la philosophie de l'esprit des différentes tentatives de naturalisation de l'intentionnalité en sciences cognitives. Notre point de départ est la question du dépassement du fonctionnalisme au cœur de ces tentatives, tel qu'on peut le voir à l'œuvre dans le débat entre le fonctionnalisme computationnel et le fonctionnalisme biologique (Millikan 1999), entre la conception modulaire de l'esprit et la conception interactive (Churchland, Ramachandran et Sejnowski 1994) ou encore entre le représentationnalisme et les théories écologiques de la cognition (Gibson 1979). Ce qui est en jeu dans ces différents débats, c'est à chaque fois la question du dépassement d'un des présupposés du fonctionnalisme computationnel lié au projet du cognitivisme classique vers la prise en compte des dimensions contextuelles et corporelles de la cognition<sup>1</sup>.

Cependant, par delà la question analytique des différentes formes de fonctionnalisme issues du projet de la naturalisation et du débat autour du choix entre l'une ou l'autre école de pensée, nous voulons montrer l'intérêt d'élargir l'interrogation vers la question de l'usage des croyances épistémiques mobilisées dans la critique du fonctionnalisme computa-

<sup>1</sup> Nous avons argumenté pour un tel dépassement dans Dedeurwaerdere (2002).

tionnel. Or dans la littérature contemporaine, on peut distinguer deux types de croyances épistémiques mobilisés dans les critiques du fonctionnalisme<sup>2</sup>. Le premier, lié au projet du naturalisme fort, considère que le projet de naturalisation de l'intentionnalité doit s'inscrire dans le même cadre épistémique que les sciences de la nature. L'accent est mis sur la conception stratifiée de la nature formant une unité emboîtée du physique à l'intentionnel (Mundale et Bechtel 1996: 483), l'unité historique de l'évolution de l'homme naturel à l'homme symbolique (Barkow, Tooby et Cosmides 1992: 5) ou encore sur la présupposition d'une réalité indépendante comme arrière-plan non intentionnel (*nonintentional background*) de toute description scientifique de l'intentionnalité (Searle 1995: 177-89). Cette première position maintient donc les présuppositions épistémiques de la position fonctionnaliste, tout en essayant de les dépasser au niveau de sa pratique de modélisation. Le second, lié au projet du naturalisme faible, envisage plus radicalement les conséquences du dépassement du fonctionnalisme sur ses propres croyances épistémiques et considère que toute théorie intentionnelle doit nécessairement rester interne à son propre usage. L'accent sera alors mis sur le dépassement de la position empirico-formelle de la modélisation à laquelle est encore lié le naturalisme fort pour évoluer vers une forme de réalisme pragmatique (Poirier 1997), sur les dimensions sémantiques et pragmatiques de la modélisation comme production intentionnelle (Putnam 1989) ou encore, de façon plus radicale, sur la constitution réflexive de l'arrière-plan non intentionnel de la modélisation comme expérience vécue d'une appartenance plus originaire au monde (Fisette et Poirier 2000: 321-23).

Toutefois, qu'il s'agisse des formes de naturalisme faible ou fort, l'usage de ces croyances épistémiques est rarement construit pour lui-même. De plus, il n'est pas clair sur quoi se base la dissociation entre ces deux formes de naturalisme et la mobilisation tantôt schématique, tantôt réflexive de l'arrière-plan de présupposés non intentionnels. Notre hypothèse est que l'usage des croyances épistémiques en sciences cognitives ne se limite pas à l'alternative entre naturalisme faible et naturalisme fort et qu'on peut envisager un autre usage de ces croyances qui reste en deçà

<sup>2</sup> Voir Fisette et Poirier (2000: 318-21). Voir également Maesschalck (2001: 237-60) pour une confrontation croisée de la conception d'arrière-plan comme structure de capacité dans la ligne de J. Searle d'une part et comme texture ouverte des croyances dans la ligne de H. Putnam d'autre part.

de cette alternative. Serait-il possible d'envisager un rapport réflexif à l'arrière-plan des présupposés non intentionnels dans le naturalisme fort ou, inversement, un rapport schématique à ces présupposés dans le naturalisme faible ? Ou encore, comment envisager le passage de la constitution des arrière-plans épistémiques, que ce soit de façon schématique ou réflexive, vers de nouvelles possibilités de modélisation ? À travers ces questions nous voulons montrer que le déplacement épistémologique qu'implique la prise en compte de l'intentionnalité en sciences cognitives, d'une forme de naturalisme fort vers un naturalisme faible, nécessite également de construire autrement l'évolution des croyances épistémiques d'arrière-plan qu'elle présuppose.

### 1 La critique du fonctionnalisme dans le naturalisme fort

Le point de départ de notre réflexion est l'insuffisance du fonctionnalisme classique ou computationnel si l'on veut élargir le programme de recherche des sciences cognitives au projet de naturalisation de l'intentionnalité. Au point de départ de ce projet de naturalisation, on trouve une certaine acception commune de l'intentionnalité, que l'on peut caractériser dans les termes de Brentano comme étant la propriété des états mentaux par laquelle ils sont dirigés vers des objets et doués d'un certain contenu (Engel 1994: 550). À la différence des objets étudiés en physique ou en biologie, la cognition humaine définit elle-même ses propres buts, les visées intentionnelles, et crée son propre champ de significations, les contenus intentionnels. C'est précisément cette double polarité de l'intentionnalité, la visée des buts, sous la forme des fonctions cognitives à réaliser, et la création des significations, par des opérations de transformation d'information, que nous retrouvons dans les modèles de l'intentionnalité en sciences cognitives.

Dans le fonctionnalisme classique on s'intéresse aux mécanismes computationnels qui caractérisent les opérations intentionnelles, que ce soit les mécanismes de calcul logique dans le cognitivisme classique ou les mécanismes d'association par similarité et d'encodage vectoriel dans le connexionnisme. Cette première perspective purement computationnelle est toutefois insuffisante si l'on veut modéliser également les aspects représentationnels et fonctionnels de l'intentionnalité. La prise en compte de la problématique de la modélisation des dimensions représentationnelles et fonctionnelles de l'intentionnalité, dans le cadre du projet fort de naturalisation, implique d'élargir le fonctionnalisme classique

dans une double direction. Tout d'abord, il faut intégrer dans la modélisation la façon dont le système cognitif construit son propre environnement opératoire ou milieu de vie. Ensuite, il faudra prendre en compte l'insertion des fonctions cognitives dans les dynamiques autonomes plus larges de cet environnement qui jouent un rôle dans leur stabilisation. Nous illustrerons cette double progression à partir de deux exemples particuliers d'élargissement du fonctionnalisme classique : le fonctionnalisme biologique inspiré des travaux de R. Millikan d'une part et le fonctionnalisme systémique inspiré des travaux de A. Clark d'autre part. Le premier tente de prendre en compte l'aspect téléologique de l'organisation autonome du contexte représentationnel des systèmes cognitifs et le second tente d'intégrer dans les modèles fonctionnalistes l'insertion des fonctions cognitives dans des dynamiques environnementales plus larges.

### 1.1 *Le fonctionnalisme computationnel classique*

Au préalable il nous faut d'abord définir de façon plus précise en quoi consiste le fonctionnalisme computationnel classique qui définit le projet des sciences cognitives. Pour ce faire, nous étudierons de plus près les réflexions méthodologiques d'un des fondateurs du fonctionnalisme classique, David Marr. La position de Marr est particulièrement intéressante pour notre propos, en ce qu'elle permet de définir une forme de fonctionnalisme qui se distingue autant des tentatives plus sophistiquées d'inspiration biologique que celles d'inspiration systémique qui seront toutes les deux au cœur du projet de la naturalisation de l'intentionnalité (Marr 1982 ; Kitcher 1988).

Les réflexions de Marr visent à préciser la spécificité de l'approche computationnelle de la cognition, à la fois par rapport à une approche purement syntaxique ou algorithmique et par rapport à une réduction biologique, à partir des neurosciences. En introduisant les distinctions entre computation, algorithme et implémentation, Marr entend critiquer aussi bien les approches purement neurophysiologiques de la vision (niveau de l'implémentation) que les approches purement syntaxiques (niveau algorithmique). Plus particulièrement, Marr a voulu se distancer par rapport à la méthodologie du programme de recherche naissant de la vision artificielle (Kitcher 1988: 2). Dans ce programme on suppose que la vision procède en deux étapes : d'abord la segmentation de l'image en traits (*features*) et ensuite le regroupement des traits dans des catégories en vue

de la reconnaissance de l'objet. Ce scénario plausible se base, entre autres, sur les fameuses expériences de Barlow, Hubel et Wiesel d'enregistrement isolé des activités neuronales. Par ces expériences on a découvert des neurones « détecteurs » au niveau du cortex visuel, c'est-à-dire des neurones s'activant uniquement quand un trait spécifique (couleur rouge, courbure anguleuse, forme quadratique, etc.) se présente dans le champ visuel (Marr 1982: 12–14). Cependant, malgré le succès initial de ce programme, on n'a jamais découvert de détecteurs de traits de plus haut niveau, correspondant au regroupement hypothétique des traits en un tout signifiant. En plus, l'implémentation sur ordinateur d'un programme de segmentation d'images en traits s'est révélée très difficile à cause de la complexité des images réelles<sup>3</sup>.

À côté des problèmes empiriques, le programme de recherche de la vision artificielle se confronte également à de sérieux problèmes d'ordre théorique. Marr soulève deux grandes objections aux présuppositions théoriques de ce programme de recherche : la première portant sur l'hypothèse de détection de traits par des neurones isolés et la deuxième sur la recherche d'un algorithme de reconnaissance. L'étude de ces deux objections nous permettra de mieux comprendre la nécessité d'introduire un niveau d'étude spécifique par rapport au niveau neurophysiologique (la détection des traits) et par rapport au niveau algorithmique (l'algorithme de reconnaissance), niveau qui sera celui du traitement d'information, appelé par Marr le niveau computationnel.

La première objection de Marr peut être illustrée à partir du problème appelé dans la littérature le problème du « neurone grand-mère ». Supposons que la vision procède effectivement par segmentation de l'information visuelle en traits et que, à partir de là, le système visuel recompose l'image en vue de la reconnaissance de l'objet. Alors, hypothétiquement, lors de la reconnaissance, le neurone ayant comme entrée l'ensemble des traits correspondants à l'objet devrait décharger. Le « neurone grand-mère », c'est le neurone qui, selon cette théorie, décharge quand je reconnais ma grand-mère (Kitcher 1988: 2–3). Supposons que l'on ait trouvé le « neurone grand-mère ». Supposons même que l'on puisse retracer la chaîne des stimuli nerveux de la rétine jusqu'à ce

<sup>3</sup> Comme le suggère par exemple le réseau de reconnaissance des visages de Cottrell, il n'y a probablement pas une segmentation de l'image en traits, mais une segmentation en composantes holistiques ou prototypes (cf. Churchland 1995: 47).

neurone. Que saurait-on alors de la façon dont nous reconnaissons des grand-mères ? Quelle information le système visuel utilise-t-il ? Comment combine-t-il cette information ? On pourrait voir le chemin des stimuli nerveux, mais on ne saurait pas pourquoi les neurones empruntent ce chemin. En bref, on ne saurait pas ce que le système visuel est en train de faire quand nous reconnaissons notre grand-mère (Kitcher 1988: 3).

Une objection similaire peut être soulevée par rapport à la recherche de l'algorithme de reconnaissance. En effet, les premiers programmes de vision artificielle se basaient sur une théorie de la vision en deux étapes : d'abord la détection des traits et ensuite, un algorithme de reconnaissance. Le but de l'algorithme est de simuler les performances du système visuel humain. En partant d'une approximation raisonnable de l'information disponible sur la rétine, le programme doit fournir en sortie approximativement les mêmes classifications perceptives que l'être humain pour un certain domaine d'application (Kitcher 1988: 3). Mais un tel algorithme ne nous apprend de nouveau rien sur ce que le système visuel est en train de faire. Il fournit seulement la structure logique ou syntaxique reliant l'entrée à la sortie du système visuel. De plus, on peut construire de nombreux algorithmes qui produisent des opérations de transformation entrée-sortie équivalentes, sans savoir lequel est effectivement utilisé par le système visuel.

Ni la détection des neurones qui s'activent lors d'une reconnaissance, ni la syntaxe des opérations du système visuel ne permettent de comprendre ce qu'est la vision. Ce qu'il faut comprendre, c'est la fonction de traitement d'information réalisée par le système visuel. On doit pouvoir spécifier la fonction de transformation des informations de l'environnement présentes au niveau de la rétine jusqu'aux représentations de l'environnement objectivement visible pour nous (Kitcher 1988: 4). À la différence des approches syntaxiques — qui étudient les algorithmes qui calculent de façon effective la fonction de transformation entrée-sortie — les contenus informationnels traités par le système cognitif ne sont pas simplement syntaxiques, mais ont une signification par rapport à la *fonction* computationnelle. Le concept de computation, entendu comme fonction de transformation d'informations, permet alors de donner une première définition du niveau de détermination propre à la pensée qui caractérise le projet des sciences cognitives dans son ensemble.

## 1.2 *Le fonctionnalisme biologique*

Dans l'approche computationnelle de Marr, les représentations manipulées par les systèmes cognitifs n'ont de signification que par rapport à la fonction de traitement d'information dans laquelle elles interviennent. Ainsi, dans un programme de reconnaissance de formes par exemple, les représentations manipulées par un tel programme — les angles, les courbes, les surfaces, etc. — n'ont une signification que par rapport à la fonction de reconnaissance de formes. Cette approche suppose donc que les représentations sont des données neutres présentées à l'entrée du système de traitement de l'information et qu'elles reçoivent leur signification uniquement de l'interprétation fonctionnelle. Par-là l'approche computationnaliste classique s'inscrit encore dans un paradigme profondément physicaliste. En effet, ni la définition de la fonction globale à réaliser par le système, ni la définition des données sur lesquelles il opère ne sont prises en compte en tant que telle dans la modélisation. Elles sont simplement présentes à titre de conditions frontières de la modélisation. Par conséquent la modélisation se limitera à l'étude des fonctions algébriques de transformation « entrée - sortie » d'un système computationnel.

Cependant une telle approche mathématique est une abstraction par rapport aux opérations cognitives d'un système intentionnel concret et ne convient donc pas à la modélisation de l'intentionnalité. En effet, comme nous l'avons vu, ce qui caractérise un système intentionnel est sa capacité sémantique à créer son propre champ de significations ou milieu de vie et sa capacité fonctionnelle à définir ses propres buts. On peut illustrer cette distinction entre le milieu de vie d'un organisme vivant et l'environnement objectif des systèmes computationnels abstraits de Marr à partir de l'exemple d'un système cognitif extrêmement simple, celui de l'abeille (Uexküll 1965: 52). On a pu montrer que les abeilles se posent de préférence sur les figures qui ont une forme ouverte, comme les étoiles et les croix, et évitent celles qui ont une forme fermée, comme le carré et le cercle. Le système perceptif de l'abeille possède donc une capacité de discrimination de formes, mais ces formes sont tout à fait spécifiques pour son espèce. Si l'on suppose par exemple comme environnement une prairie en fleurs, dans laquelle alternent les fleurs écloses et les boutons, le milieu de vie correspondant sera alors simplement donné par l'ensemble de traits perceptifs, comme les étoiles et les cercles. Ces traits

perceptifs spécifiques du milieu servent les intérêts biologiques de l'abeille : seules les fleurs (étoiles et croix) et non les boutons (cercles) ont une signification pour l'abeille.

Si d'un point de vue théorique les deux catégories de représentation véridiques ou relatives à un milieu de vie sont possibles, des travaux récents suggèrent que l'encodage des représentations dans un format spécifique à l'action est omniprésent et fondamental. Supposons par exemple que l'on voie un léopard à la chasse. On peut se demander comment le système visuel fait correspondre les points individuels d'une image du léopard en mouvement, observé à un certain moment, par exemple une perception des taches sur son dos, aux points de l'image qui succède immédiatement. Il existe une solution computationnelle élégante à ce problème, proposée par Shimon Ullman, un collaborateur de David Marr (Ullman 1979). Ullman suggère que le système visuel calcule différentes combinaisons possibles des points et, parmi celles-ci, prend la combinaison qui produit la distance minimale globale entre les points des deux images. Mais ce calcul exhaustif du mouvement de chaque point n'est absolument pas nécessaire, vu que dans des situations normales les taches du léopard ne vont pas bouger dans des sens différents. On peut supposer qu'il s'agit d'un mouvement d'ensemble et, à partir de la correspondance d'une seule tache d'une image à l'autre, extrapoler la correspondance directement aux autres points. C'est probablement ce que fait le système visuel, comme a pu le montrer de façon convaincante V. S. Ramachandran dans ses expériences sur les effets d'entraînement de mouvement (Ramachandran 1990). Dans la perception du mouvement du léopard, le système visuel néglige des quantités énormes d'information de la scène visuelle, concernant par exemple le détail du mouvement des points individuels, et construit, en combinant des informations partielles, une représentation utile de la scène (Ramachandran 1990: 348–49). Le raccourci de calcul ne risque pas de nous induire en erreur, si on applique cette stratégie de calcul dans des contextes appropriés, par exemple chaque fois que l'on rencontre des animaux tachetés en mouvement dans notre milieu de vie.

Cette importance de l'usage des représentations partielles et circonstanciées par les systèmes cognitifs aura des conséquences sur la définition de la notion de représentation dans les processus cognitifs. En effet, dans ces modèles d'inspiration biologique, les représentations ne sont pas uniquement des données qui interviennent dans les fonctions mathématiques

de traitement d'information, mais sont des états internes qui sont intégrés dans l'organisation fonctionnelle du système nerveux. D'après l'analyse de Andy Clark on peut caractériser ces représentations au sens biologique par la combinaison de trois conditions (Clark 1997: 143–47) :

- (a) Un système sera un système représentationnel si on peut identifier des états internes qui ont comme fonction spécifique de véhiculer une information sur des états de choses corporels ou environnementaux,
- (b) si la relation de représentation entre les états internes et les états de choses est systématique, elle définit un schéma d'encodage, et
- (c) si le schéma d'encodage est suffisamment général ou complexe, portant sur un ensemble large de stimuli venant de l'environnement ou du corps.

Selon cette caractérisation, un état interne sera une représentation s'il a comme fonction de véhiculer une information suffisamment générale sur ce qui se passe dans l'environnement. La simple présence d'un schéma d'encodage (condition b), c'est-à-dire d'une corrélation causale systématique entre l'état interne et un trait du corps ou de l'environnement, ne suffit pas pour parler d'une représentation. Ainsi, même s'il existe une corrélation causale entre les marées et la lune, nous ne dirons pas que la mer représente la lune ou inversement. En effet il ne nous paraît pas plausible que les marées (par exemple) aient été sélectionnées, conçues ou aient évolué dans le but de véhiculer une information sur la position de la lune. Ce qui importe, en plus de la corrélation causale, c'est le rôle que joue l'état interne dans le système dont il fait partie. Il faut que ce rôle soit précisément de véhiculer de l'information et que cette information soit effectivement utilisée par d'autres états, c'est-à-dire qu'il y ait d'autres états qui « consomment » l'information (condition a). Toutefois, même si le système a été conçu ou a évolué pour véhiculer l'information de son environnement, nous n'avons pas encore une condition suffisamment spécifique pour parler d'une représentation. Un tournesol qui se tourne en fonction de la position du soleil est corrélé de façon causale avec la position du soleil et cette corrélation a une fonction adaptative pour le tournesol. Toutefois nous ne dirons pas que le tournesol représente la position du soleil. De même un simple robot, conçu pour réagir à

une source lumineuse et se mouvoir vers cette source de façon autonome, détecte la lumière mais ne construit pas pour autant des représentations. Il faut de plus que la corrélation causale soit suffisamment complexe : « Parler en termes de représentation commence à faire sens si l'on est en présence d'états internes qui sont corrélés de façon systématique à tout un ensemble de contingences de l'environnement » (Clark 1997: 147 ; notre traduction).

Pour illustrer cette définition sur un exemple concret, on peut considérer les représentations de bas niveau, comme les représentations de la position des membres du corps. Chez le rat par exemple, les neurones du cortex pariétal postérieur représentent l'information concernant les directions de mouvement et les opérations du réseau de neurones les catégorisent en gauche, droite et tout droit (Clark 1997: 144). Malgré une corrélation causale assez directe avec les positions du corps, il s'agit bien d'une représentation : nous avons un schéma d'encodage général (chaque catégorie de direction couvre différents stimuli) et le système nerveux du rat « consomme » l'information sur la direction pour aider le rat à courir, comme on peut l'observer à partir d'expériences d'enregistrement électrique de l'activité des neurones du cortex du rat (Clark 1997: 146).

Cette définition des représentations dans la ligne des travaux de Ruth Millikan, John Haugeland et de Andy Clark permet de préciser le déplacement par rapport à la position fonctionnaliste qu'implique la définition par le système cognitif de leur propre environnement opératoire ou milieu de vie. Contrairement à l'analyse de Marr, on ne peut plus dire que la fonction détermine entièrement la signification des représentations. En effet, comme nous l'avons vu, du point de vue de l'implémentation, la représentation n'est pas seulement définie dans son rapport au traitement de l'information, mais également par un schéma d'encodage dans le cerveau. En termes techniques, la représentation suppose un schéma d'encodage général qui définit des opérations de transformation logiques des stimuli captés par le système nerveux en état représentationnel interne (la condition b et c de l'analyse de Clark). Du point de vue de l'implémentation, c'est donc l'encodage qui crée le contexte sémantique et pas la fonction. Si l'on veut prendre en compte cette autonomie du contexte sémantique dans la modélisation de l'intentionnalité, il faudra donc croiser la perspective fonctionnelle avec une étude de la cohérence ou de la topologie interne du champ des significations.

### 1.3 Le fonctionnalisme systémique

Le croisement entre la perspective computationnelle et la perspective biologique déterminant une auto-organisation des significations en fonction des intérêts vitaux d'un organisme permet de définir une approche de modélisation qui prend en compte la définition par un système intentionnel de son propre champ de significations. Cependant, afin de construire une opération de modélisation spécifique de l'intentionnalité, il faut également considérer la capacité qu'ont les systèmes intentionnels à définir leurs propres buts à l'intérieur d'un environnement qui possède sa propre autoprogrammation. Dans cette perspective, l'environnement n'a pas seulement un rôle passif dans la réalisation des fonctions cognitives, mais il peut également jouer un rôle actif dans le processus d'ajustement entre les activités d'un agent cognitif et son environnement opératoire.

Pour donner une première idée de l'interaction des systèmes cognitifs avec les ressources dynamiques de leur environnement, on peut reprendre un exemple donné par Andy Clark dans *Being There* (Clark 1997: xii). Considérons le cas d'une tâche manuelle, où l'on doit emboîter des pièces finement ajustées. C'est une tâche que l'on essaie de simuler en robotique pour construire des lignes d'assemblage entièrement automatisées. D'après Clark on peut envisager deux solutions très différentes à ce problème de simulation. La première solution suggérée par Clark est la solution mentaliste, qui ne tient pas compte de l'interaction avec l'environnement : « Confronté à la question du contrôle d'un robot par un ordinateur pour réaliser une tâche d'emboîtement de pièces, on peut construire de multiples boucles de rétroaction. Ces boucles peuvent signaler à l'ordinateur si le robot n'a pas réussi à emboîter une pièce et lui permettre de recalculer le mouvement dans une direction légèrement différente » (Clark 1997: xii ; notre traduction). Toutefois, on peut également envisager une deuxième solution, qui tient compte des ressources dynamiques de l'environnement : « La cognition corporelle et contextuelle aborde le problème d'une façon différente. On peut par exemple attacher simplement le bras du robot à un joint en caoutchouc, qui laisse du jeu dans deux directions de l'espace. Dans cette nouvelle situation, l'ordinateur ne doit plus calculer les boucles de régulation fine du mouvement. Les pièces 'se glissent et se fauillent' en bonne position comme si de multiples ajustements de rétroaction étaient calculés continuellement » (Clark

1997: xii ; notre traduction). Dans cette deuxième solution, le joint en caoutchouc remplace le calcul des ajustements par l'ordinateur.

Des expériences en psychologie du développement suggèrent que pour la coordination du mouvement des membres du corps, le cerveau exploite également de telles boucles de rétroaction élargies. Par exemple, dans un modèle développé par Esther Thelen et Linda Smith (Thelen et Smith 1994), la coordination du mouvement des membres s'appuie sur des dynamiques corporelles intrinsèques, comme l'élasticité des muscles et la raideur des membres, et des éléments environnementaux, comme la force gravitationnelle. Il n'y a pas une exécution de commandes motrices internes qui spécifient une trajectoire détaillée de mouvement, mais simplement la modulation de différents facteurs qui interviennent dans le mouvement. Si les dynamiques corporelles jouent un rôle si important, la modulation devrait être différente d'une personne à l'autre. Dans une expérimentation de Thelen et Smith sur le développement du comportement de préhension chez les enfants, on constate en effet une telle différence d'un enfant à l'autre. Un enfant, Gabriel, était très actif par nature. Il générait des mouvements rapides du bras. Pour lui la difficulté était de convertir les battements en un mouvement dirigé. Pour ce faire il devait apprendre à contracter les muscles au moment où son bras s'approchait d'un objet cible, de façon à ralentir le battement et à permettre un contact approprié. Hannah, au contraire, était calme par nature. Les mouvements qu'elle produisait étaient limités. Son problème n'était pas de contrôler les battements, mais de générer suffisamment d'élan pour vaincre la gravité (Thelen et Smith 1994: 247-78). Dans ces expériences, chaque enfant partait d'un mélange de dynamiques corporelles intrinsèques différent. Même si la tâche — atteindre un objet — était la même pour tous les enfants, le problème de coordination à résoudre pour chaque enfant se présentait sous une forme différente. L'apprentissage des nouveaux comportements est donc autant conditionné par le développement de schèmes mentaux de contrôle que par les dynamiques corporelles intrinsèques.

On se demandera peut-être quelle peut être la pertinence de ces modèles systémiques inspirés de la cybernétique pour les sciences cognitives. Est-ce qu'on n'a pas tout simplement changé d'objet d'étude, de la cognition vers la régulation ? Et s'il n'y a plus de représentations, ni de computations, peut-on encore parler d'opérations *cognitives* ? Tout d'abord les modèles cybernétiques permettent de mieux circonscrire l'usage — plus limité qu'on ne le croyait — des représentations. Parfois, comme dans le

cas de la balle à attraper, un couplage dynamique avec l'environnement permet de résoudre un problème que classiquement on résout de façon computationnelle. Le choix entre une solution cybernétique et une solution représentationnelle, comme dans le cas du choix entre les représentations véridiques et les représentations partielles, dépendra du problème auquel on est confronté. Mais l'intérêt des modèles cybernétiques va plus loin. En effet dans la plupart des situations on rencontre un mélange de solutions cybernétiques et de solutions représentationnelles. Dans les systèmes cognitifs complexes, comme les nôtres, les interactions avec l'environnement et les computations sur des représentations coopèrent de façon permanente.

La prise en compte du couplage dynamique des systèmes cognitifs avec leur environnement opératoire permettra de compléter les perspectives ouvertes par le fonctionnalisme biologique. Par conséquent l'étude de l'intentionnalité, telle qu'on peut la définir dans le cadre du projet fort de naturalisation, supposera non seulement de dépasser le computationnalisme classique de Marr, comme c'était déjà le cas dans le fonctionnalisme d'inspiration biologique, mais également de croiser ce fonctionnalisme biologique avec l'étude de la stabilisation des systèmes en couplage avec leur environnement. À partir de ce croisement, on pourra alors envisager une pratique de modélisation particulière, qui permet de rendre compte des dimensions représentationnelles et fonctionnelles de l'intentionnalité au sein même de la pratique de la modélisation.

## 2 La critique du fonctionnalisme dans le naturalisme faible

Dans son ouvrage *Being there*, Andy Clark écrit que la solution des problèmes conceptuels en sciences cognitives dépend ultimement du progrès empirique qui sera réalisé dans la discipline (Clark 1997: 175 et 168-69). Si l'on accepte ce point de vue, caractéristique du naturalisme fort, alors les concepts n'interviennent plus qu'à titre d'heuristiques, d'images ou de métaphores d'un programme de recherche entièrement naturalisé, sans pouvoir être justifiés pour eux-mêmes. La lumière nouvelle que le naturalisme fort entend jeter sur le rapport entre la représentation et l'action dans les modèles de l'intelligence reste donc muet sur l'importance des choix conceptuels qui président aux différentes positions particulières qui sont défendues dans la discipline. C'est pourquoi une deuxième forme de naturalisme, qu'on peut qualifier de faible, critique le présupposé du réalisme scientifique qui préside au naturalisme fort, pour mettre en évi-



dence la problématique des choix conceptuels opérés dans l'élaboration des modèles particuliers de l'intentionnalité.

### 2.1 *La critique du réalisme scientifique dans le naturalisme faible*

Pour comprendre l'origine de la relativité conceptuelle des modèles de l'intentionnalité en sciences cognitives, il faut comprendre les effets sur le statut de la modélisation de la prise en compte de la dimension représentationnelle et de la dimension fonctionnelle de l'intentionnalité. Nous aborderons d'abord la question de la spécificité des modèles fonctionnels, en rapport avec le concept de système téléonomique. Ensuite nous montrerons quelle modification supplémentaire il faut introduire, quand on tient compte également de la dimension représentationnelle dans la modélisation.

Dans la première section, nous avons étudié la prise en compte de la dimension fonctionnelle dans le cadre des conceptions de la cybernétique. Les systèmes cybernétiques sont des systèmes qui tendent vers un certain but selon la dynamique de l'interaction avec l'environnement, interaction qui est définie par des boucles de rétroaction. Cependant, comme le montre Frederic Suppe à partir d'une analyse des lois fonctionnelles, l'étude de la seule dynamique de l'interaction ne suffit pas pour connaître le but vers lequel tendra le système (Suppe 1989: 162–65). Ce but dépend également des conditions de stabilisation de la dynamique d'interaction, conditions qui sont réalisées par la présence de certains éléments stables dans l'environnement avec lequel le système interagit.

Un exemple intéressant de l'importance des conditions de stabilisation des systèmes cybernétiques est donné par la théorie de la sélection naturelle en biologie. D'après la théorie de la sélection naturelle, la compétition des espèces dans un écosystème entraîne la disparition des espèces les moins adaptées et la survie des espèces les plus adaptées. Toutefois cette adaptation dépend de l'environnement dans lequel on opère. Si par exemple on modifie artificiellement les sources de nutrition dans l'écosystème, on pourra obtenir soit une augmentation, soit une diminution de la biodiversité dans le système. L'état d'équilibre vers lequel tend le système dépend des contraintes présentes dans l'environnement. Cet exemple illustre bien que, dans l'étude des systèmes cybernétiques, les éléments de l'environnement avec lesquels le système interagit font partie

du modèle. De plus, comme le précise bien Suppe, « pour un certain but donné, spécifié par une fonction, [...] la définition des conditions frontières qui portent sur les paramètres de stabilisation dans l'environnement implique une nouvelle loi d'évolution du système » (Suppe 1989: 165 ; notre traduction). Dans le cas des systèmes cybernétiques, les lois d'évolution du système varient selon les contraintes présentes dans l'environnement.

C'est précisément l'importance des conditions de stabilisation dans la modélisation en sciences cognitives que nous avons pu étudier dans le paragraphe de la première section consacré au fonctionnalisme systémique. Cependant, si l'on veut tenir compte également de la dimension représentationnelle, en plus de la dimension fonctionnelle, l'analyse de Suppe doit être modifiée. En effet, dans le cas des systèmes cognitifs, le contexte de fonctionnement n'est pas l'environnement naturel mais un contexte représentationnel ou un milieu de vie. Comme nous l'avons vu, l'environnement des systèmes cognitifs est un environnement représentationnel construit à partir de schémas d'encodage généraux implémentés dans le cerveau. Un même environnement stabilisateur sera donc représenté d'une façon différente selon la niche sémantique du système cognitif en question. Par conséquent, les effets du contexte stabilisateur sur le système cognitif seront également différents selon le contexte représentationnel en question.

Cette interdépendance entre les modèles du contexte fonctionnel et les modèles du contexte représentationnel implique que la position du réalisme scientifique qui caractérise les sciences de la nature n'est plus tenable si l'on veut tenir compte des dimensions représentationnelles et fonctionnelles dans la modélisation de l'intentionnalité. Dans les sciences de la nature, les modèles particuliers sont construits à partir des principes généraux en spécifiant les hypothèses auxiliaires qui permettent de donner une valeur particulière aux variables et aux paramètres du modèle (Bunge 1973: 30). Ces hypothèses auxiliaires portent sur la définition des conditions frontières et des conditions initiales du système (Suppe 1989: 71). À la différence des sciences de la nature cependant, la définition des modèles particuliers de l'intentionnalité en sciences cognitives ne dépend pas uniquement de conditions qui portent sur la frontière entre le système et l'environnement, mais dépend également d'un modèle du contexte opératoire du système cognitif, à la fois dans sa dimension fonctionnelle

et représentationnelle. Par conséquent, les lois d'évolution du système seront différentes selon les modèles de ce contexte.

Ainsi, par exemple, dans les expériences sur la coordination motrice chez le jeune enfant, on peut montrer que les modèles de coordination du mouvement seront très différents selon les dynamiques corporelles intrinsèques sur lesquelles s'appuie le système de coordination. En effet le but poursuivi par le système de régulation sera différent selon que le problème auquel l'enfant est confronté est d'atténuer la dynamique de son corps ou, au contraire, d'engendrer suffisamment d'élan pour réaliser un certain mouvement. Les opérations cognitives optimales qui permettent de réaliser cette fonction seront donc différentes selon le modèle du couplage fonctionnel avec l'environnement. De façon général, si l'on prend en compte la dimension représentationnelle et fonctionnelle des opérations intentionnelles dans la pratique de la modélisation, les lois d'évolution du système seront différentes selon les hypothèses auxiliaires qui spécifient ce contexte opératoire.

## 2.2 *La relativité conceptuelle des modèles*

Nous pouvons tirer deux conclusions de notre discussion de l'importance de la prise en compte du contexte dans l'étude de l'intentionnalité. Premièrement, les lois d'évolution du système seront différentes selon le modèle du contexte. Deuxièmement, il y a une interdépendance entre les modèles du contexte fonctionnel et les modèles du contexte représentationnel : le rôle du système de représentation varie en fonction du contexte fonctionnel (modification de la dynamique de stabilisation) et les effets du contexte fonctionnel dépendent du système de représentation (modification de la signification du contexte).

Cette interdépendance des facteurs contextuels permet de construire des interprétations alternatives d'une même expérience. Par exemple, dans l'expérience de la coordination motrice chez l'enfant, nous avons fait varier les modèles du couplage fonctionnel, mais nous avons supposé le modèle du système de représentation du corps constant. En l'occurrence, dans les deux modèles du couplage fonctionnel, le système cognitif représente uniquement les aspects du corps qui interviennent dans la modulation du mouvement. Autrement dit, la représentation du corps est une représentation pratico-sociale. Toutefois on pourrait construire une autre interprétation de la même expérience en modifiant à la fois le modèle du contexte représentationnel et le modèle du couplage.

Ainsi, l'expérience de la coordination motrice chez l'enfant peut être modélisée également à partir d'un modèle mentaliste de contrôle centralisé où le rôle du corps se réduit à fournir des informations à un système de représentation exhaustive du mouvement.

Le choix entre différentes interprétations d'une même expérience n'est évidemment pas arbitraire. Il devra s'appuyer sur des hypothèses auxiliaires qui permettent de justifier tel ou tel modèle du contexte représentationnel et fonctionnel. Toutefois, en l'absence d'hypothèses auxiliaires bien établies, le choix entre les modèles ne sera plus déterminé par des critères scientifiques seuls. Dans ce cas, le choix entre les modèles sera fonction d'une interprétation de ce qu'est l'intentionnalité.

Remarquons que cette problématique de l'interprétation est souvent occultée dans la recherche en sciences cognitives. En effet, comme dans les exemples de la coordination motrice et de la représentation de l'espace chez l'enfant, on s'intéresse soit à la dimension fonctionnelle, soit à la dimension représentationnelle du contexte. Le modèle de l'autre dimension est présupposé à titre d'arrière-fond de la modélisation. C'est précisément en explicitant cet arrière-fond que l'on peut se rendre compte du facteur interprétatif qui guide la modélisation.

Afin d'illustrer l'importance de ce rôle de l'interprétation dans la construction des modèles particuliers en sciences cognitives, on peut reprendre l'exemple de l'utilisation de représentations partielles et circonstanciées dans le système visuel. Ainsi, comme nous l'avons vu, pour construire une représentation d'une surface de points mobiles, on peut inférer le mouvement des points individuels à partir du mouvement d'ensemble, sans devoir représenter le mouvement de chaque point individuel de façon explicite. Toutefois, même si une telle solution est adéquate dans des contextes particuliers, elle pourrait tomber en défaut si on était confronté à une surface où tous les points ne se meuvent pas de façon solidaire. Pour éviter ce problème, on peut partir de modèles plus classiques, où l'on suppose la construction de représentations exhaustives de l'environnement.

Ramachandran justifie son choix pour les représentations partielles en invoquant un critère d'utilité auquel doivent satisfaire les représentations : « On pourrait argumenter, plutôt, que la perception est essentiellement un "stock d'astuces" [...] qui n'ont pas été adoptées pour des raisons d'attrait esthétique ou d'élégance mathématique, mais simplement parce qu'elles marchaient (d'où le terme de théorie "utilitariste" de la

perception) » (Ramachandran 1990: 347; notre traduction). Dans l'interprétation de Ramachandran, la perception est essentiellement un guide d'action efficace, par rapport à des tâches spécifiques. Ramachandran argumente pour la plausibilité de cette position en invoquant des arguments tirés de la théorie de l'évolution et des arguments computationnels. Les premiers arguments visent à montrer l'importance du couplage des systèmes cognitifs à un environnement de vie. Les deuxièmes veulent établir la performance d'un système algorithmique basé sur des dispositifs de calculs circonstanciés.

Toutefois ces arguments sont simplement des hypothèses auxiliaires de sa modélisation et ne justifient pas entièrement le choix pour l'interprétation utilitariste. Une interprétation alternative plus classique, qui attribue au système perceptif la tâche de construire des représentations exhaustives de l'environnement, reste également plausible. Les hypothèses auxiliaires qui pourraient appuyer une telle hypothèse mettraient davantage l'accent sur l'indépendance du système par rapport à l'environnement, plutôt que sur le couplage, et sur l'importance des capacités d'abstraction des systèmes cognitifs, plutôt que sur leur flexibilité contextuelle.

Cet exemple de conflit d'interprétation en sciences cognitives illustre l'importance des critères non scientifiques qui interviennent dans la modélisation de l'intentionnalité. Le champ des modèles particuliers de l'intentionnalité ne peut pas être spécifié de façon univoque sans l'intervention d'hypothèses auxiliaires, venant d'autres disciplines, qui s'appuient sur une certaine interprétation de ce qu'est l'intentionnalité. L'analyse de l'élaboration des modèles particuliers des opérations intentionnelles montre donc non seulement la nécessité, pour un formalisme donné, de spécifier le modèle du contexte fonctionnel et représentationnel, mais également que le domaine d'application peut avoir une signification contextuelle propre, indépendante du formalisme, déterminée par une interprétation de ce qu'est l'intentionnalité.

### 3 La problématique de l'évolution des croyances épistémologiques

La critique du fonctionnalisme dans le naturalisme faible semble donc imposer une refonte radicale du cadre épistémologique des sciences cognitives. En effet, à la différence des critiques du fonctionnalisme dans le

naturalisme fort, la critique dans le naturalisme faible met en cause la possibilité de définir une approche fonctionnaliste alternative de l'intentionnalité, comme par exemple à partir d'un croisement des propositions du fonctionnalisme biologique et du fonctionnalisme systémique. Plutôt, la déconstruction de certaines notions du fonctionnalisme classique dans le projet de la naturalisation, comme la notion de représentation ou la notion d'environnement comme source de données, permet de mettre en évidence un déplacement épistémologique qui montre la relativité conceptuelle des modèles de l'intentionnalité en sciences cognitives. Cette hypothèse de travail rejoint le projet de Hilary Putnam dans *Représentation et réalité*, où il montre les insuffisances des différentes améliorations des modèles fonctionnalistes de l'intentionnalité en sciences cognitives quand ils continuent à adopter le cadre épistémologique du réalisme scientifique qui caractérise les sciences de la nature. C'est ce qui amène Putnam à formuler l'hypothèse épistémologique alternative d'un réalisme interne, qui devrait pouvoir mieux prendre en compte la spécificité du domaine de l'action intentionnelle. Sur les traces de H. Putnam (1989/1990: 113-16), nous pouvons formuler à titre préliminaire l'hypothèse d'un tel déplacement vers le réalisme interne par une conditionnalité à deux niveaux :

- condition d'autoréférentialité : une théorie qui entend modéliser l'intentionnalité doit pouvoir s'interroger elle-même comme production intentionnelle ;
- condition d'usage : l'application des modèles de l'intentionnalité sera conditionnée par une certaine interprétation de ces modèles interne à leur usage dans le contexte concret de l'action.

Ces deux conditions épistémologiques renvoient le projet de la modélisation de l'intentionnalité à la question des conceptions de l'action en amont de la modélisation d'une part et à son application au domaine de l'action de l'autre.

Même si les travaux empiriques discutés dans cette contribution ne sont pas les mêmes que ceux considérés par Putnam à l'époque de *Représentation et réalité*, il nous semble que cette double conditionnalité du réalisme interne reste tout à fait pertinente pour dépasser les insuffisances épistémologiques de la modélisation de l'intentionnalité dans le cadre fonctionnaliste. En effet, c'est précisément cette double condition que nous avons mobilisée dans notre critique du fonctionnalisme dans la

perspective du naturalisme faible. Tout d'abord, la condition d'autoréférentialité suppose une réinterprétation du statut de la pratique scientifique, pour être comprise elle-même comme une production intentionnelle à part entière. C'est précisément une réinterprétation de la pratique scientifique dans le sens à prendre en compte ses propres dimensions sémantiques et contextuelles qui a permis de mettre en évidence le rôle des conditions de stabilisation et de contextualisation dans l'élaboration des modèles particuliers de l'intentionnalité. Ensuite, au niveau de la condition d'usage, nous avons vu que l'équilibre particulier entre la perception contextualisante du contexte et les conditions de stabilisation du contexte n'est pas déterminé par des critères scientifiques seuls, mais par une pratique interprétative de ce qu'est l'intentionnalité.

La critique radicale du fonctionnalisme dans le naturalisme faible repose donc sur un nouvel ensemble de croyances épistémologiques par rapport au cadre du réalisme scientifique du naturalisme fort. Cependant cette évolution dans l'arrière-plan des croyances épistémologiques du réalisme scientifique vers un réalisme interne est rarement élucidée pour elle-même. Est-ce qu'elle est une conséquence automatique du changement d'objet d'étude, passant du domaine des objets du monde naturel vers les objets intentionnels, ou est-ce qu'une certaine forme de réalisme scientifique est néanmoins encore compatible avec la modélisation de l'intentionnalité comme le prétend par exemple le naturalisme fort ? De plus, supposons qu'on admette que la naturalisation de l'intentionnalité implique une évolution dans les croyances épistémologiques, qu'advient-il des programmes fonctionnalistes novateurs du naturalisme fort qui entendaient également dépasser le cadre trop restreint du fonctionnalisme mathématique ? Autrement dit, est-ce qu'une élucidation de la question du passage du naturalisme fort vers le naturalisme faible pourrait également donner des indications en retour sur la signification du fonctionnalisme dans le projet fort de naturalisation ?

### 3.1 *L'incomplétude des croyances épistémologiques mobilisées dans le naturalisme fort*

Afin d'élucider la question de l'évolution des croyances épistémologiques et, corrélativement, la question d'un retour éventuel sur la signification du projet fonctionnaliste, il nous faut d'abord préciser le statut des croyances épistémologiques dans le naturalisme fort. À ce titre, nous

partirons des conceptions développées par John Searle dans son livre *The Construction of Social Reality* où il développe un projet qui est proche de notre entreprise. En effet, dans ce livre, il développe une conception de l'intentionnalité qui prend en compte les dimensions représentationnelles et fonctionnelles de l'intentionnalité, tout en procédant à une élucidation du statut épistémologique de l'explication fonctionnelle causale du comportement intentionnel. Toutefois, à la différence des arguments que nous avons avancés en faveur du réalisme interne, il continue à défendre la position du réalisme scientifique. Sa position devrait donc nous permettre d'explicitier l'arrière-plan des croyances épistémologiques mobilisées — souvent de façon implicite — par le naturalisme fort.

Tout d'abord, dans l'analyse de Searle, il y a une dépendance entre la justification du réalisme scientifique d'une part et la théorie de l'intentionnalité de l'autre. Comme l'explique Searle, la justification du réalisme scientifique dépend de la théorie de l'intentionnalité en ce que le réalisme scientifique fonctionne comme une présupposition d'arrière-plan de la pratique scientifique comme production intentionnelle. En effet on ne peut pas baser la justification du réalisme scientifique sur cette pratique elle-même, vu qu'alors la justification présupposerait déjà ce qu'il y a à justifier. Néanmoins, sans une telle présupposition, cette pratique cesserait simplement d'être intelligible. Comme l'écrit Searle, « quand nous comprenons un énoncé du type que nous avons considéré [i.e. de type scientifique], nous le comprenons comme un énoncé qui présuppose une réalité d'accès publique » (Searle 1995/1998: 187) et ceci indépendamment du fait que « nos affirmations sont des objets de connaissance ou non, qu'elles sont vraies ou fausses et même du fait que les objets auxquels on réfère existent ou non » (Searle 1995/1998: 187).

Le projet de la naturalisation de l'intentionnalité dans le cadre du programme fort suppose donc de construire l'objet intentionnel d'une manière ou d'une autre comme faisant partie d'une réalité indépendante. Si l'on applique cette présupposition aux différentes descriptions fonctionnalistes de l'intentionnalité que nous avons étudiées dans le naturalisme fort, il faut alors identifier la référence à un tel arrière-plan dans les tentatives de modélisation de l'intentionnalité. En effet les tentatives d'intégration du fonctionnalisme systémique et du fonctionnalisme biologique, dans le but de rendre compte des propriétés de l'intentionnalité, n'est intelligible que si il se base sur une réalité externe permettant d'opérer une telle intégration. Or c'est effectivement une telle référence

que l'on trouve articulée chez la plupart des défenseurs du modèle intégré, que ce soit sous le mode spatial d'une stratification des niveaux physique, biologique et intentionnel (Mundale et Bechtel 1996 ; Bechtel 1994), sous le mode temporel d'une séquence articulant l'histoire longue d'une adaptation fonctionnelle à une séquence courte de réaction d'une teleofonction à un stimulus externe (Millikan 1984) ou sous le mode d'une référence à des concepts transversaux comme chez Andy Clark à partir des concepts de couplage, de computation et de représentation (Clark 1997: 143-76).

Toutefois par la dissociation entre les mécanismes causaux qu'on peut décrire dans un cadre fonctionnaliste d'une part et l'analyse purement formelle des arrière-plans représentationnels de l'autre, le naturalisme fort occulte les hypothèses auxiliaires qui interviennent dans l'application des cadres formels à des modèles causaux particuliers. En effet cette analyse purement formelle de l'arrière-plan interprétatif ne permet pas de spécifier les interprétations de l'intentionnalité qui seront effectivement mobilisées dans tel ou tel modèle particulier. Or, du point de vue de la définition des modèles particuliers de l'intentionnalité, nous avons vu que le domaine de l'action intentionnelle n'est pas un objet de référence « en général » d'une opération de formalisation, mais que sa spécification dépend d'une pratique interprétative qui définit un équilibre particulier entre la perception contextualisante du contexte d'une part et les conditions contextuelles de stabilisation de cette perception de l'autre. Cette spécification implique donc un déplacement d'un cadre de modélisation de l'intentionnalité comme objet général d'une représentation scientifique vers des modèles de l'intentionnalité entendue comme action en contexte. Par ce déplacement, qu'on trouve à l'œuvre dans le naturalisme faible, on peut alors rendre compte de l'association au sein des situations concrètes de la causalité mécaniste et de la causalité intentionnelle que le naturalisme fort avait dissociées.

### 3.2 *L'incomplétude des croyances épistémologiques mobilisées dans le naturalisme faible*

Néanmoins, le projet de la naturalisation faible de l'intentionnalité, tel qu'il se construit dans le cadre du réalisme interne, mobilise également un ensemble de croyances épistémologiques qui ne peuvent pas être justifiées sur une seule base empirique. En effet, les modèles particuliers de

l'intentionnalité ne permettent pas de stabiliser l'équilibre entre le contexte fonctionnel et le contexte représentationnel, dans la mesure où ces modèles admettent, comme nous l'avons vu, une pluralité de pratiques interprétatives en cohérence avec des hypothèses auxiliaires de modélisation. La position alternative du réalisme interne, telle que nous l'avons définie ci-dessus dans la suite des travaux de Putnam à partir de la double conditionnalité, opère donc une idéalisation des conditions d'acceptation des modèles qui risque de reproduire les insuffisances épistémologiques du réalisme scientifique si elle n'éclaire pas la question de l'accès à ces conditions idéalisées. En effet la pratique interprétative permettant de définir ces conditions fournit un critère d'adéquation des modèles sous des conditions épistémiques suffisamment bonnes (Putnam 1981: 54). Toutefois, comme l'écrit Putnam dans un ouvrage récent sur la question du réalisme, on peut se poser la question de savoir « comment l'on peut avoir un accès référentiel ou autre aux "conditions épistémologiques suffisamment bonnes" » (Putnam 1999: 18 ; notre traduction). Une telle image reste encore prisonnière des présuppositions épistémologiques de la modernité en ce qu'elle « retient la prémisse d'une interface entre le sujet connaissant et tout le reste "à l'extérieur" » (Putnam 1999: 18 ; notre traduction). La relativité conceptuelle des modèles, tant qu'elle est comprise comme un schème mental qui s'entrepasse entre nos modèles et une réalité externe à modéliser, ne pourra pas fournir un critère de stabilisation de la pratique de modélisation, étant donné que l'on n'a pas d'accès référentiel à ces hypothèses auxiliaires de la modélisation.

Cependant la stabilisation s'opère bel et bien et c'est ce qui amène Putnam à envisager, à la suite de philosophes comme W. James, E. Husserl, L. Wittgenstein et J. Austin (Putnam 1999: 24), la position alternative du « réalisme naturel » que l'on peut également appeler un réalisme non fondationnel : « notre croyance dans le réalisme ne tient pas au fait qu'une réalité externe à laquelle nous n'avons pas accès "cause" des expériences mentales internes, selon l'hypothèse des *qualia*, mais tient au fait que les objets "externes", choux et rois, peuvent être objet d'expérience » (Putnam 1999: 20 ; notre traduction). Cette hypothèse alternative d'un lien génétique entre la croyance dans une réalité externe qui opère dans le réalisme interne et celle qui opère dans l'expérience naturelle du monde permet d'interpréter autrement la stabilisation du projet de naturalisation. En effet, en nous basant sur la proposition de Putnam, nous pouvons concevoir le troisième terme, qui permet de stabi-

liser l'ajustement entre la perception contextualisante et les conditions contextuelles de stabilisation de la perception, de façon plus adéquate comme une disposition cognitive à la stabilisation (Maesschalck 2001: 206), une ouverture au remplissage possible de la règle par le monde, telle qu'elle opère déjà dans nos « transactions » (Maesschalck 2001: 169) naturelles avec les objets de ce monde. Précisons que ce que Putnam appelle le réalisme naturel ne peut aucunement être confondu avec le réalisme du sens commun. Il s'agit bien d'une posture réflexive interne au jeu de langage philosophique, une « seconde naïveté » (Maesschalck 2001: 44) comme le dit Putnam, qui ne vise pas à proposer une nouvelle théorie de la perception ou une « nouvelle image métaphysique » (Maesschalck 2001: 41 ; traduction de l'auteur), mais une reconstruction réflexive de « l'accès conceptuel que nous avons aux choses au sujet duquel nous parlons et réfléchissons » (Maesschalck 2001: 45 et 157 ; traduction de l'auteur)<sup>4</sup>.

### 3.3 Les conséquences du déplacement épistémologique

Quelles conclusions peut-on tirer de cette reconstruction systématique des croyances épistémologiques mobilisées dans le naturalisme fort et le naturalisme faible ? Et quelles sont les implications d'une compréhension réflexive de l'arrière-plan des croyances épistémologiques ?

Tout d'abord les développements qui précèdent permettent de mettre en évidence une double insuffisance dans la construction des croyances épistémologiques dans le projet de naturalisation. Cette insuffisance se marque par l'incomplétude des présuppositions d'arrière-plan qui tente de fixer les pratiques de modélisation par la production d'un schème d'unité de cette pratique.

Dans le cas du naturalisme fort, l'avant-plan de la pratique de modélisation est constitué par le domaine de l'action intentionnelle, dont essaient de rendre compte les différentes tentatives fonctionnalistes de croisement de la dimension représentationnelle et fonctionnelle du contexte de l'action intentionnelle. Comme nous l'avons vu à partir de l'analyse de J. Searle, cet avant-plan dépend de la présupposition de l'existence d'une

réalité indépendante qui conditionne l'intelligibilité de ces tentatives de croisement. Cette présupposition permet de comprendre le domaine de l'action intentionnelle comme une réalité indépendante. L'accent est donc mis sur le schème d'unité de l'objet intentionnel permettant d'articuler l'autotranscendance de la pratique de modélisation, comme condition d'intelligibilité formelle de celle-ci. Cependant, comme nous l'avons vu, le renvoi à cette autotranscendance se fait au prix d'une dissociation formelle entre l'ordre causal décrit par les modèles fonctionnalistes et l'ordre intentionnel déterminant les conditions d'intelligibilité.

Dans le cas du naturalisme faible, l'avant-plan est constitué par l'unité d'une méthode de modélisation de l'intentionnalité qui reste interne à l'usage des modèles de l'intentionnalité. Cette compréhension de la modélisation, que ce soit dans une conception sémantique de la pratique de modélisation (Suppe 1989) ou dans une conception plus pragmatique (Fisette et Poirier 2000), permet de mettre en évidence la relativité conceptuelle des modèles, ou encore, la pratique interprétative de l'intentionnalité qui permet de stabiliser les modèles fonctionnalistes particuliers. Le naturalisme faible permet donc de rétablir l'association entre l'ordre causal mécanique et l'ordre causal intentionnel. Néanmoins cette association se fait sous un mode internaliste au dépend de l'abandon de la visée d'autotranscendance. De fait, comme le met bien en évidence la critique du dernier Putnam, l'idéalisation des conditions épistémiques fait l'impasse sur la question de l'accès référentiel ou autre à ces conditions épistémiques.

Ce qu'il faudrait donc envisager, afin de dépasser cette double insuffisance, c'est un usage des croyances épistémologiques qui permet à la fois de maintenir l'autotranscendance des pratiques du naturalisme fort et l'internalisme de l'association de l'ordre intentionnel et causal du naturalisme faible. L'enjeu ne semble donc pas tellement être le choix entre l'alternative du naturalisme fort et le naturalisme faible, que l'élaboration d'une position épistémologique intermédiaire qui permet de sauvegarder les avantages des deux positions, sans pour autant reproduire le procédé de généralisation schématique qui fige l'arrière-plan des croyances épistémologiques.

Cette conclusion de notre confrontation croisée des croyances épistémologiques mobilisées dans le naturalisme fort et le naturalisme faible permet également de considérer autrement la proposition d'une voie différente qui reconstruit l'arrière-plan épistémologique comme une posture

<sup>4</sup> Pour les conséquences ultérieures de ce modèle de la disposition cognitive sur une compréhension réflexive de l'opération de contextualisation comme opérateur de modalisation, voir Maesschalck (1998).

réflexive à l'intérieur de la pratique de la modélisation. Une telle voie alternative, que nous avons considérée dans la ligne des derniers travaux de Putnam, permet effectivement de dépasser l'incomplétude du naturalisme faible<sup>5</sup>. En effet elle met en évidence le lien génétique entre la disposition réflexive, qui détermine l'équilibre entre l'intentionnalité et son contexte, et la disposition cognitive à l'autotranscendance qui opère déjà dans l'évidence naturelle du monde. Cependant, d'après ce qui précède, le modèle de la réflexivité ne pourrait effectivement dépasser la double insuffisance que si elle envisage en même temps sa traduction dans des formes opératoires d'unité réflexive pratique qui associent de façon internaliste l'ordre causal et l'ordre intentionnel dans la pratique de la modélisation<sup>6</sup>. Dans une telle perspective, on pourrait par exemple envisager un autre rôle des théories interchamps mobilisées dans le naturalisme fort. Plutôt qu'un rôle schématique, qui réfère de façon formelle à une unité d'action transversale déjà donnée dans les conditions générales de l'expérience, ces théories pourraient avoir un rôle prospectif et viser, sous un mode constructiviste, la constitution d'enjeux transversaux pour l'action qui peuvent mener à leur tour vers l'émergence de nouvelles pratiques interprétatives dans l'élaboration des modèles particuliers de l'intentionnalité.

## Références

- Barkow, J. H., J. Tooby et L. Cosmides (1992). *The Adapted Mind*. New York : Oxford University Press.
- Bechtel, W. (1994). « Biological and Social Constraints on Cognitive Processes. The Need for Dynamical Interactions Between Levels of Organization ». *Canadian Journal of Philosophy*, Supplementary Volume 20, pp. 133–164.
- Blakemore, C., dir. (1990). *Vision : Coding and Efficiency*. Cambridge : Cambridge University Press.
- Bunge, M. (1973). *Method, Model and Matter*. Boston/Dordrecht : D. Reidel.

<sup>5</sup> On trouve une telle position également dans les travaux sur l'intentionnalité opérante dans la phénoménologie husserlienne des années vingt (Husserl 1998) ou encore dans les travaux sur le réalisme direct (Smith 1995), pour autant qu'on n'interprète pas ces travaux dans un sens ontologique mais comme une posture réflexive interne à notre commerce quotidien avec le monde. Sur ce dernier point, voir également Fisette et Poirier (2000: 321–23).

<sup>6</sup> C'est une telle réflexivité « étendue » à sa traduction contextuelle que tente de développer pour lui-même la recherche menée par Lenoble et Maesschalck (2003).

- Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge, MA : MIT Press.
- Churchland, P. S., V. S. Ramachandran, et T. J. Sejnowski (1994). « A Critique of Pure Vision ». In Koch et Davis (1994), pp. 23–60.
- Clark, A. (1997). *Being There : Putting Brain, Body and World Together Again*. Cambridge, MA/London : MIT Press.
- Cummins, R., A. Ariew et M. Perlman, dir. (2002). *Functions: New Readings in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press.
- Dedeurwaerdere, T. (2002). *Action et Contexte. Du tournant cognitiviste à la phénoménologie transcendantale*. Hildesheim / Zürich / New York : Olms.
- Engel, P. (1994). « La philosophie de l'esprit ». In Meyer (1994), pp. 529–64.
- Fisette, D. et P. Poirier. (2000). *Philosophie de l'esprit. Etat des lieux*. Paris : Vrin.
- Gibson, J.J. (1979). *The Ecological Approach to Visual perception*. Boston : Houghton Mifflin.
- Husserl, E. (1998). *De la synthèse passive. Logique transcendantale et constitutions originaires*. Trad. par Bruce Bégout et Jean Kessler avec la collaboration de Natalie Depraz et Marc Richir. Grenoble : Jérôme Millon.
- Kitcher, P. (1988). « Marr's Computational Theory of Vision ». *Philosophy of Science* LL, pp. 1–24.
- Koch, C. et J. Davis, dir. (1994). *Large-scale Neuronal theories of the Brain*. Cambridge, MA : MIT Press.
- Lenoble, J. et M. Maesschalck (2003). *Toward a Theory of Governance. The Action of Norms*. London / New York : Kluwer Academic Publishers.
- Maesschalck, M. (1998). *Typique transcendantale et typique phénoménologique, I. Les enjeux d'une recherche sur la typique de la raison pure*. Les carnets du Centre de Philosophie du Droit LX. Louvain : Université catholique de Louvain.
- Maesschalck, M. (2001). *Normes et Contextes. Les fondements d'une pragmatique contextuelle*. Hildesheim / Zürich / New York : Olms.
- Marr, D. (1982). *Vision*. San Francisco : Freeman.
- Meyer, M., dir. (1994). *La philosophie anglo-saxonne*. Paris : Presses Universitaires de France.
- Millikan, R. (1984). *Language, Thought and Other Biological Categories. New Foundations for Realism*. Cambridge, MA : MIT Press.
- Millikan, R. (1999). « Wings, Spoons, Pills and Quills : A Pluralist Theory of Functions ». *Journal of Philosophy* XCVI(4), pp. 191–206.
- Mundale, J. et Bechtel W. (1996). « Integrating Neuroscience, Psychology and Evolutionary Biology through a Teleological Conception of Function ». *Mind and Machines* VI, pp. 481–505.

- Poirier, P. (1997). *Explication scientifique et théorie représentationnelle de l'esprit : enquête sur une théorie représentationnelle de l'esprit*. Thèse de doctorat. Montréal : Université du Québec à Montréal.
- Putnam, H. (1981). *Reason, Truth and History*. Cambridge : Cambridge University Press.
- Putnam, H. (1989). *Representation and Reality*. Cambridge, MA : MIT Press.  
Trad. par C. Engel-Tiercelin (1990). *Représentation et réalité*. Paris : Éditions Gallimard.
- Putnam, H. (1999). *The Threefold Cord : Mind, Body and World*. New York : Columbia University Press.
- Ramachandran, V. S. (1990). « Interactions between Motion, Depth, Color and Form : The Utilitarian Theory of Perception ». In Blakemore (1990), pp. 346–60.
- Searle, J. (1995). *The Construction of Social Reality*. New York : Free Press.  
Trad. par C. Tiercelin (1998). *La construction de la réalité sociale*. Paris : Éditions Gallimard.
- Smith, B. (1995). « Formal Ontology, Common Sense and Cognitive Science ». *International Journal of Human-computer Studies* XLIII, pp. 641–67.
- Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Chicago : University of Illinois Press.
- Thelen, E. et L. Smith (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA : MIT Press.
- Uexküll, J. V. (1965). *Mondes animaux et monde humain*. Paris : Denoël.
- Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA : MIT Press.