

Pacherie, E. 1995. Attitudes propositionnelles, intentionnalité et évolution. *Revue de Métaphysique et de Morale*, 100, 3: 339-371.

## **ATTITUDES PROPOSITIONNELLES, INTENTIONNALITE ET EVOLUTION**

Elisabeth PACHERIE

La question du statut des attributions d'attitudes propositionnelles ou, sous une autre formulation, celle de la relation entre la psychologie ordinaire et la psychologie scientifique que les sciences cognitives aspirent à édifier, constitue l'un des principaux problèmes actuellement débattus en philosophie de l'esprit et de la psychologie. Cette question entretient des liens étroits avec certaines problèmes philosophiques traditionnels, comme celui du rapport du corps et de l'esprit, celui de l'intentionnalité, celui de la nature d'une explication scientifique, celui de la nature de la rationalité, qui trouvent dans ces débats des formulations et une acuité nouvelles. A travers les réponses qu'elle suscite, elle sert aussi de révélateur à certains désaccords fondamentaux.

De la grande variété des réponses émergent deux courants principaux. Le premier courant, réaliste, prend ses racines philosophiques chez les penseurs de l'époque classique, aussi bien empiristes comme Locke et Hume que rationalistes comme Descartes. Les tenants du réalisme sont mentalistes et tendent à reconnaître aux concepts de la psychologie ordinaire le statut d'objets théoriques de plein droit. Ils se donnent pour but la formulation d'une théorie psychologique scientifique qui se conforme aux exigences du monisme matérialiste tout en préservant au moins dans une large mesure les intuitions de la psychologie ordinaire.

Le second courant est naturaliste et éliminativiste et s'inscrit dans la lignée de penseurs comme Quine et Dewey. Le courant éliminativiste tient la psychologie ordinaire pour une théorie inadéquate dont les concepts sont des artefacts sans contrepartie objective et qui doit être éliminée et remplacée. Ce rejet est motivé par des considérations de nature épistémologique et ontologique. Néanmoins, à la suite

de Quine et en vertu du principe de la double norme dont il s'était fait l'avocat<sup>1</sup>, certains des éliminativistes contemporains, tels Stich et Dennett, tout en refusant un statut théorique et une réalité ontologique aux attitudes propositionnelles, admettent l'utilité prédictive des attributions d'attitudes propositionnelles. Éliminativisme théorique et "conservatisme" pragmatique se trouvent ainsi conjugués. Néanmoins, les éliminativistes ne peuvent se contenter de prendre acte des succès prédictifs de la psychologie des attitudes propositionnelles. Encore faut-il qu'ils expliquent pourquoi et comment la psychologie des attitudes propositionnelles peut avoir de tels succès prédictifs si, comme ils le soutiennent, elle n'est pas vraie.

Dennett s'est pendant longtemps lui-même qualifié d'éliminativiste matérialiste et d'instrumentaliste et s'est efforcé de montrer que scepticisme ontologique à l'égard des attitudes propositionnelles et croyance en la valeur heuristique des attributions d'attitudes propositionnelles étaient deux positions que l'on était fondé à adopter conjointement. Dans son livre de 1987, *The Intentional Stance*, Dennett apporte certaines nuances aux positions qu'il avait défendues. Il y rejette l'étiquette d'instrumentaliste au profit de celle de réaliste modéré. Il mobilise de nouveaux arguments pour expliquer le succès prédictif de la psychologie des attitudes propositionnelles et fait en particulier appel à la théorie de la sélection naturelle. Dans ses arguments darwiniens, Dennett attribue à la théorie de la sélection naturelle le rôle de théorie-relais entre une psychologie scientifique qui n'admet pas les croyances et les désirs au sein de son ontologie et une théorie de l'interprétation qui leur conserve une valeur prédictive.

Le but de cet article est de proposer une analyse critique des arguments darwiniens de Dennett. Pour mesurer la portée des enjeux liés à ces arguments, il importe de connaître les raisons qui motivent son scepticisme vis-à-vis des thèses réalistes et l'alternative qu'il suggère dans sa théorie des systèmes intentionnels. Dans un premier temps, je présenterai donc les grandes lignes de l'approche réaliste

---

<sup>1</sup> Dans la section 45 de son livre *Le mot et la Chose*, Quine définit une norme ontologique à laquelle doit se plier une théorie scientifique et une norme pragmatique qui régit notre pratique usuelle. Ainsi:

Si nous voulons dépeindre la structure véritable et ultime de la réalité, le schème canonique qui nous convient est le schème austère, qui ne connaît pas d'autres citations que la citation en discours indirect, et qui ne connaît pas d'attitudes propositionnelles, mais seulement la constitution de la matière et le comportement des organismes. [...] Mais si nous nous servons de la notation canonique simplement pour dissoudre des perplexités verbales ou rendre plus aisées les déductions logiques, nous serons souvent bien avisés de tolérer les constructions de l'attitude propositionnelle." (Quine, 1960: trad. Gochet, p. 307)

et les raisons de son rejet par Dennett. J'examinerai ensuite la théorie des systèmes intentionnels et les objections qu'à son tour elle suscite. Enfin, j'analyserai les arguments darwiniens qu'utilise Dennett pour contrer ces objections et expliquerai pourquoi ils me semblent peu convaincants.

## **1. Réalisme, fonctionnalisme et computationnalisme**

Nous expliquons spontanément notre comportement et celui d'autrui en termes de croyances, désirs et autres états représentationnels (craintes, souhaits, intentions, espérances, etc.) que nous nous attribuons à nous-mêmes et aux autres. Ces états représentationnels que les philosophes ont regroupés depuis Russell sous l'appellation générale d'attitudes propositionnelles appartiennent au stock de concepts que la psychologie ordinaire utilise pour décrire, expliquer ou prédire les comportements. Les explications psychologiques ordinaires semblent véhiculer un certain nombre de présupposés dont les principaux sont les suivants:

- (1) Les attitudes propositionnelles sont des états mentaux dotés d'une existence objective.
- (2) Ces états mentaux interagissent causalement tant entre eux qu'avec des entités physiques.
- (3) Ces états mentaux ont un contenu et des conditions de satisfaction. Autrement dit, ce sont des états intentionnels.
- (4) Le rôle causal de ces états mentaux est déterminé par leur contenu.

Le réalisme en matière d'attitudes propositionnelles consiste à prendre au sérieux les présupposés de la psychologie ordinaire et les généralisations qu'elle propose. Ainsi en est-il de Fodor qui donne du réalisme la définition suivante:

Je propose de dire que quelqu'un est un Réaliste au sujet des attitudes propositionnelles si et seulement si (a) il soutient qu'il y a des états mentaux dont les occurrences et interactions causent le comportement et le font, qui plus est, de manière à respecter (tout au moins dans une large mesure) les généralisations de la psychologie des désirs et des croyances du sens commun; et (b) il soutient que ces mêmes états mentaux causalement efficaces sont également sémantiquement évaluables. (Fodor, 1985: 78).

Une théorie psychologique réaliste des états mentaux se donne donc pour but d'expliquer comment il peut y avoir des états mentaux qui aient les propriétés sémantiques et causales qu'attribue aux attitudes propositionnelles la psychologie ordinaire. Cette explication doit être scientifiquement recevable, ce qui présuppose qu'elle soit compatible avec le monisme matérialiste.

Les réalistes contemporains adhèrent à ce que Fodor a appelé la théorie représentationnelle de l'esprit: ils admettent l'existence de représentations mentales dotées d'un contenu sémantique et d'une efficacité causale. En cela, ils sont les continuateurs d'une tradition philosophique qui a ses racines tant dans l'empirisme que dans le rationalisme classique. Ce qui distingue les défenseurs contemporains de la théorie réaliste de leurs prédécesseurs est qu'ils souscrivent très majoritairement au monisme matérialiste. Ils s'efforcent pour la plupart de concilier monisme matérialiste et réalisme en matière d'attitudes propositionnelles via l'adhésion à l'approche computationnelle et au fonctionnalisme.

Le paradigme computationnel postule que les représentations mentales sont de nature symbolique et sont manipulées selon des règles formelles. Le fonctionnalisme, dans sa version la plus couramment acceptée, se définit par les deux thèses suivantes: (1) chaque état mental particulier se range dans une certaine catégorie ou type d'état mental en fonction du rôle qu'il joue dans l'explication psychologique, c'est-à-dire en fonction de ses relations à des entrées, des sorties et d'autres états mentaux; (2) chaque état mental particulier est identique à un état neurophysiologique, mais il n'y a pas nécessairement correspondance biunivoque entre type d'état mental et type d'état neurophysiologique. Par sa première thèse, le fonctionnalisme se distingue du béhaviorisme qui entend définir les états mentaux par leurs relations à des entrées et sorties mais refuse de prendre en compte ses relations avec d'autres états mentaux. Par sa deuxième thèse, le fonctionnaliste adhère à la théorie de l'identité occasionnelle du physique et du mental et refuse la théorie plus forte de l'identité des types. Le fonctionnalisme s'efforce ainsi de concilier deux exigences: celle de l'autonomie d'un niveau d'explication psychologique et celle d'une explication matérialiste de l'efficacité causale du mental.

Selon la théorie de l'identité occasionnelle du mental et du physique, c'est en vertu du fait qu'ils sont chacun identique à un état physique particulier que les états mentaux peuvent être causalement agissants. C'est au niveau de réalisation physique du mental que se produit l'efficacité causale. Cela ne signifie pourtant pas qu'un état mental, soit, en tant qu'il est mental, causalement inerte. Le fonctionnalisme va donc de pair avec le computationnalisme lorsqu'il exploite l'idée de système formel développée depuis Frege par les logiciens et mathématiciens (Russell, Hilbert, Tarski, Gödel, Church, Turing). La notion de système formel met en évidence la possibilité d'une double articulation, reliant d'une part la syntaxe et la sémantique, d'autre part la syntaxe et la physique.

Un système formel est un système tel que les transformations auxquelles les éléments du système sont soumis sont déterminées uniquement par la structure

morphologique de ces éléments (leur forme) et les règles qui régissent chaque type de structure morphologique. L'élaboration d'une notion précise de système formel a montré la possibilité de construire des systèmes déductifs formels dont les opérations préservent la valeur sémantique qu'une interprétation peut indépendamment attribuer aux éléments du système (Théorème de complétude de Gödel). En d'autres termes, la syntaxe simule la sémantique.

La seconde articulation reliant cette fois syntaxe et physique est réalisée par "l'incarnation" du calcul inférentiel dans un calculateur matériel. On appelle machine de Turing un système formel automatique qui peut être caractérisé abstraitement mais également réalisé physiquement. Les travaux sur la calculabilité de Church et Turing ont montré qu'à chaque fonction calculable on peut faire correspondre une machine de Turing particulière, la calculant automatiquement. Autrement dit, il a été démontré qu'il est possible de modéliser les traits syntaxiques d'un système par des propriétés physiques particulières de constituants d'une machine réelle, de manière à ce que les lois de transition de la machine telles qu'elles s'expriment au niveau physique coïncident avec les règles de la syntaxe. En outre, Church et Turing ont démontré l'existence de machines de Turing dites universelles, c'est-à-dire capables d'effectuer tout calcul effectuable par une machine de Turing particulière.

Dans ce double parallélisme, la syntaxe joue, on le voit, un rôle de médiation capital puisqu'elle permet d'associer propriétés sémantiques et propriétés causales d'un symbole. L'élaboration d'une notion précise de système formel et de calculabilité et leur exploitation par les ordinateurs, constituent aux yeux des réalistes une avancée décisive pour les sciences cognitives. Ainsi, Fodor proclame-t-il que:

Les ordinateurs apportent une solution au problème de la médiation entre les propriétés causales des symboles et leurs propriétés sémantiques. Par conséquent, *si* l'esprit est une sorte d'ordinateur, nous commençons à voir comment avoir une théorie du mental qui réussisse là où littéralement toutes les tentatives précédentes n'avaient essuyé qu'échecs cuisants, une théorie qui explique comment il pourrait y avoir des relations de contenu non-arbitraires entre pensées causalement reliées. (Fodor, 1987: 19).

L'optimisme qui transparaît dans ces lignes doit cependant être nuancé. Le point fort du fonctionnalisme turingien est qu'il paraît fournir une explication de l'efficacité causale des états mentaux acceptable d'un point de vue matérialiste. En revanche, on peut douter qu'il permette de pleinement rendre compte de la sémantité des états mentaux. La métaphore de l'ordinateur montre comment il est possible de simuler des processus sémantiques par des processus formels, mais elle

ne permet pas de comprendre comment les symboles formels sont interprétés, comment ils peuvent renvoyer à une réalité extérieure au système. Le système n'interprète pas lui-même les symboles qu'il manipule, ce sont le concepteur et plus tard l'utilisateur, tous deux extérieurs au système, qui projettent sur les symboles une interprétation. On voit donc difficilement comment on pourrait accorder à un système informatique une intentionnalité intrinsèque. Son intentionnalité apparaît comme une intentionnalité d'emprunt, dérivée de celle du concepteur et de l'utilisateur.

## **2. Austérité ontologique et théorie des systèmes intentionnels**

Ce n'est cependant pas cette faiblesse du réalisme qui en éloigne Dennett. Dennett ne voit pas dans le problème de l'intentionnalité intrinsèque un problème philosophique profond mais au contraire un faux-problème, un artefact qui a sa source dans l'intempérance ontologique des réalistes. Selon lui, l'intentionnalité intrinsèque est un mythe et toute tentative d'en faire la théorie est par principe vouée à l'échec.

Contre l'intempérance ontologique des réalistes dont le fonctionnalisme turingien est à ses yeux une manifestation, Dennett prêche avec Quine l'austérité ontologique. Le fonctionnalisme turingien est selon lui une thèse beaucoup trop forte. L'idée que chaque type d'état mental est identique à un type d'état fonctionnel susceptible d'être décrit comme un état logique d'une machine de Turing n'est pas plus crédible que l'idée selon laquelle à chaque type d'état mental devrait correspondre un type d'état physique. L'erreur des réalistes consiste à prendre trop au sérieux l'idiome intentionnel et à s'obstiner à postuler pour la plupart des termes de la psychologie ordinaire l'existence de référents objectifs. Lorsque nous attribuons à autrui une croyance, cette attribution nous permet un certain nombre de prédictions sur son comportement. Mais que nous puissions tirer un pouvoir prédictif de l'attribution d'états intentionnels n'implique nullement que nous devons prendre ces attributions littéralement et admettre que les croyances ont leur place dans le "meuble du monde physique".

Le scepticisme ontologique de Dennett trouve une de ses sources dans son adhésion à la thèse de l'indétermination radicale de la traduction de Quine. Cette thèse peut être étendue, comme Quine lui-même le souligne, au cas de l'interprétation radicale, c'est-à-dire au cas où un observateur attribue à un autre individu des croyances et des désirs:

S'attendre à ce qu'il y ait un mécanisme physique distinct derrière chaque état mental authentiquement distinct est une chose; s'attendre à ce qu'il y ait

un mécanisme distinct pour chaque prétendue distinction exprimable dans le langage mentaliste traditionnel est une autre chose. La question de savoir [...] si l'étranger croit A ou croit B réellement est une question dont je me demande si elle a vraiment un sens. C'est ce que je veux suggérer quand je défends la thèse de l'indétermination de la traduction. (Quine, 1970: 180-181).

Selon Quine, il est possible qu'une entreprise de traduction radicale aboutisse à des manuels de traduction reposant sur des systèmes d'hypothèses analytiques différents qui soient tous compatibles avec la totalité des faits observables sur le comportement verbal, mais tels que les traductions proposées pour certains énoncés soient incompatibles entre elles. Il y a alors indétermination radicale de la traduction dans la mesure où toutes les traductions reposent sur des hypothèses compatibles avec la totalité des faits observables et où il n'existe pas de fait caché supplémentaire qui puisse les départager. La thèse de l'indétermination radicale se distingue de la thèse de la sous-détermination des théories scientifiques relativement aux observations, dans la mesure où cette sous-détermination pourrait être levée si des faits encore ignorés étaient connus, tandis que l'indétermination radicale persiste alors même que la totalité des faits est connue.

Pour Dennett comme pour Quine, cette interprétation vaut également pour l'interprétation radicale: "il est toujours possible en principe que des interprétations faites du point de vue intentionnel soient à égalité, en sorte qu'il n'y aurait pas de fait déterminant qui pourrait fixer ce que le système intentionnel en question croit réellement" (Dennett, 1987, trad. fr. 1990: 57). Inversement, on peut dans certaines circonstances attribuer la même croyance à plusieurs individus sans avoir à postuler qu'à cette croyance correspond dans la tête de chacun un même état fonctionnel. Selon l'exemple de Dennett (1987, tr. fr. 1990: 76), si Jacques tue son oncle d'un coup de fusil dans Trafalgar Square, si Sherlock le prend sur le fait et si Tom apprend la nouvelle dans le *Guardian* et Boris dans la *Pravda*, Jacques, Sherlock, Tom et Boris, malgré des expériences remarquablement différentes auront une chose en commun: tous croiront qu'un Français a commis un crime dans Trafalgar Square. Mais c'est, au yeux de Dennett, faire une hypothèse gratuite que de postuler l'existence d'un objet structurellement similaire dans la tête de chacun. Les seuls faits pertinents, que ce soit dans le domaine de la traduction, celui de l'interprétation ou les sciences en général sont les faits observables. De même que rien ne nous autorise à poser l'existence de significations autonomes pour briser l'indétermination de la traduction, rien ne nous autorise à postuler l'existence d'états mentaux inaccessibles à l'observation mais dont le contenu serait parfaitement déterminé.

Au fonctionnalisme turingien, qui est solidaire de l'hypothèse réaliste dans la mesure où à chaque type d'état mental il fait correspondre un type fonctionnel déterminé, Dennett propose de substituer un fonctionnalisme moins contraignant, qu'il nomme fonctionnalisme occasionnel (token-functionalism) et selon lequel "si chaque état mental est bien un état physique donné et en outre un état fonctionnel donné, les types états mentaux ne sont pas définissables comme des types d'une machine de Turing" (Dennett, 1981, xvi-xvii).

Mais qu'est-ce qui permet alors d'expliquer que dans certains cas nous puissions attribuer à deux individus la même croyance (qu'il pleut dehors) et sur cette base prédire correctement leur comportement (ils se munissent d'un parapluie)? Selon le fonctionnalisme occasionnel, attribuer à deux individus des croyances du même type, ne signifie pas leur attribuer des états fonctionnels de même type. Le fonctionnalisme occasionnel ne nous donne donc pas les moyens d'expliquer comment ces croyances peuvent causer un comportement identique.

C'est ici que Dennett fait intervenir la distinction entre *illata* et *abstracta*. A la suite de Reichenbach, Dennett admet que les termes théoriques se distinguent en deux catégories selon le type de référent qu'ils peuvent avoir. Des termes théoriques comme atome ou quark renvoient à des *illata*, des entités théoriques dont on peut inférer l'existence objective. En revanche, des termes comme centre de gravité ou équateur ne renvoient à aucune entité objective, leurs référents sont des *abstracta*, des constructions logiques ou mathématiques, utiles au calcul. Selon Dennett, croyances et désirs sont également des *abstracta*, les individus ont des croyances et des désirs de la même manière qu'ils ont des centres de gravité. Dennett s'affirme réaliste modéré et non instrumentaliste parce qu'il ne considère pas que les énoncés où il est fait référence à des *abstracta* sont des énoncés faux ou dépourvus de valeur de vérité. Selon lui, il est vrai de dire que la Lune a un centre de gravité ou que nous avons des croyances, mais il s'agit d'une vérité *cum grano salis* et non d'une vérité à prendre littéralement. C'est ce grain de sel supplémentaire qui fait de lui un réaliste modéré et non un Réaliste majuscule.

L'analogie avec les centres de gravité permet de comprendre comment Dennett peut accorder une valeur prédictive aux attributions de croyances et désirs sans accorder à ces croyances et désirs une efficacité causale. Pas plus que les centres de gravité ne sont causes de phénomènes physiques, les croyances et désirs ne sont causes du comportement. Dire que le centre de gravité d'une voiture de course est plus bas que celui d'une voiture de tourisme nous permet d'expliquer, entre autres choses, que la voiture de course ait une meilleure tenue de route, mais le centre de gravité n'est pas lui-même la cause ou une des causes de la bonne tenue de route. Ce qui cause la bonne tenue de route est le fait que l'attraction gravitationnelle de la



Terre exerce une force sur le véhicule. Mais cette force est exercée en *tous* points du véhicule et non simplement sur son centre de gravité. La résultante de toutes ces forces, leur somme vectorielle agit à travers un point, le centre de gravité, mais c'est là une chose différente.

Notons cependant que ce qui nous autorise à parler de centres de gravité n'est pas simplement l'utilité prédictive de la notion, c'est l'existence d'objets matériels, de forces d'attraction s'exerçant sur eux, de lois physiques permettant de calculer les points d'application de ces forces. La notion de centre de gravité s'insère dans une théorie physique et est légitimée par elle. A quelle théorie psychologique, à quel analogue du principe d'attraction universelle, Dennett peut-il recourir pour justifier le rôle d'*abstracta* qu'il attribue aux croyances et aux désirs?

A ces questions Dennett répond en proposant la théorie des systèmes intentionnels. L'utilisation de prédicats mentalistes sera légitimée si l'on peut énoncer des règles systématiques pour l'attribution de désirs et croyances et la prédiction de comportements et donner une justification des règles posées. La théorie des systèmes intentionnels se donnera donc pour but de "fournir une réduction du mental - un domaine dont les frontières sont au mieux fixées par la reconnaissance mutuelle et l'intuition partagée - à un domaine clairement défini d'entités dont les principes d'organisations sont familiers, relativement formels et systématiques." (Dennett, 1981: xviii).

Qu'est-ce exactement qu'un système intentionnel? Dennett définit un système intentionnel comme un système dont le comportement peut-être expliqué et prédit par l'attribution au système de croyances et désirs. Mais cette attribution à un système d'une intentionnalité n'est justifiée que relativement aux stratégies d'un agent qui cherche à justifier et prédire son comportement. Pour faire comprendre la notion de système intentionnel, Dennett introduit une distinction entre trois perspectives ou points de vue complémentaires selon lesquels un système peut être considéré.

Tout d'abord un système peut être considéré du point de vue de la conception (design stance), c'est-à-dire du point de vue des fonctions qu'il est censé remplir. Le point de vue de la conception est solidaire du fonctionnalisme occasionnel. La description fonctionnelle d'un système s'effectue de manière descendante. On commence par caractériser les tâches principales qu'un système est supposé remplir et on fait correspondre à chaque tâche un sous-système. La tâche attribuée à chaque sous-système peut à son tour être décomposée en sous-tâches qu'effectuent des sous-systèmes plus simples. La décomposition se poursuit jusqu'à ce que l'on aboutisse à des tâches extrêmement simples susceptibles d'être réalisées par des mécanismes élémentaires. L'adoption de la perspective de la conception permet

d'exorciser le spectre du fantôme dans la machine et la menace concomitante de la régression à l'infini (l'oeil interne qui regarde l'image fournie par la perception et qui a lui-même besoin d'un autre oeil pour regarder l'image qu'à son tour il fournit et ainsi de suite) Elle permet de comprendre comment des tâches qui semblent demander de l'intelligence peuvent en fin de compte être réalisées par des mécanismes simples. Au dernier niveau de décomposition seuls opèrent des mécanismes élémentaires et c'est l'organisation fonctionnelle hiérarchique qui rend compte de ce que cet ensemble d'opérations élémentaires peut en fin de compte donner lieu à un comportement global intelligent.

Outre le point de vue de la conception, on peut adopter le point de vue physique qui est celui de la réalisation matérielle du système. Deux systèmes peuvent parfaitement être identiques du point de vue de la conception tout en différant complètement du point de vue de leur réalisation physique. Une entreprise peut, par exemple, informatiser son système de comptabilité sans en changer les règles: les opérations qui étaient auparavant réalisées par des comptables munis de crayons et de papier, le sont maintenant par des circuits électroniques de l'ordinateur.

Dans le cas de systèmes simples, l'adoption du point de vue de la conception ou du point de vue physique peut permettre de prédire le comportement du système, mais plus le système est complexe moins ces stratégies ont de chance de succès. Ainsi, les programmes de jeux d'échec sont devenus tellement sophistiqués que ni le point de vue physique ni le point de vue de la conception ne permettent de prédire leur comportement au cours d'une partie. Même pour le concepteur d'un tel système, le meilleur moyen d'essayer de battre sa créature est de supposer qu'elle jouera les coups les plus rationnels étant donné les règles et les buts des échecs. Mais cela revient à adopter le troisième point de vue, le point de vue intentionnel, et à supposer que l'adversaire a le désir de gagner, qu'il a des croyances quant au but du jeu, aux règles de déplacement des pièces, à la valeur relative de telle ou telle position, qu'il a des objectifs, des stratégies pour y parvenir et ainsi de suite.

L'exemple du programme de jeu d'échec met en lumière la nature et les conditions de l'adoption du point de vue intentionnel. En premier lieu, l'exemple montre que les raisons qui justifient l'adoption de la stratégie intentionnelle sont d'ordre pragmatique: c'est sa seule utilité prédictive qui motive l'attribution de croyances et désirs. Par conséquent, l'adoption de la stratégie intentionnelle ne commande aucun engagement ontologique quant à l'existence effective des désirs et croyances. Dire qu'en début de partie l'ordinateur a déplacé son fou pour protéger sa reine et que plus tard il a sacrifié un pion dans le même but n'implique pas que l'on pose l'existence d'un état fonctionnel du système définissable comme "désir de protéger la reine". Mais il est néanmoins nécessaire que le succès des prédictions

effectuées en adoptant le point de vue intentionnel soit susceptible d'une explication du point de vue de la conception. L'exemple du jeu d'échecs nous montre également que l'adoption de la perspective intentionnelle exige que l'on fasse l'hypothèse non seulement que le système fonctionne conformément à la manière dont il a été conçu mais encore que la conception est optimale, autrement dit commande un comportement rationnel.

En poursuivant maintenant le parallélisme entre centres de gravité et croyances, nous pouvons donner une interprétation plus précise du réalisme modéré dont se réclame Dennett et de ses conséquences. Selon Dennett, croyances et désirs sont, au même titre que les centres de gravité, des *abstracta* et non des entités dotées d'une existence objective. Mais, évidemment, nous ne sommes pas maîtres de poser des *abstracta* simplement parce que tel est notre bon plaisir. De même que l'utilisation de la notion de centre de gravité *qua abstractum* doit être légitimée par son insertion dans une théorie physique, l'utilisation du vocabulaire des croyances et des désirs doit être légitimée par son insertion dans une théorie psychologique. La théorie des systèmes intentionnels et le principe de rationalité sont ainsi amenés à jouer vis-à-vis des croyances et désirs le rôle que la physique newtonienne et l'attraction universelle jouent vis-à-vis des centres de gravité.

Notons que jusqu'à présent nous n'avons fait que reculer pour mieux sauter: nous nous sommes transporté du problème de la légitimité de l'usage d'une notion vers celui de la légitimité de la théorie dans laquelle elle s'insère. Dire que ce qui légitime la théorie newtonienne ou la théorie des systèmes intentionnels est leur seule utilité prédictive reviendrait à embrasser un instrumentalisme pur et dur, ce dont Dennett se défend. Dans le cas des centres de gravité et de la théorie newtonienne, dépasser l'instrumentalisme revient à poser l'existence d'objets physiques dotés de propriétés et relations physiques existant objectivement et soumis à des forces existant objectivement. Les centres de gravité demeurent bien des *abstracta*, mais l'existence d'*abstracta* doit être réductible à celle d'entités et propriétés dotés d'une existence objective. Autrement dit, puisque pour définir les centres de gravité, il faut faire appel à l'attraction universelle, le réalisme demande qu'au principe d'attraction universelle on fasse correspondre des forces physiques objectives.

Dans la théorie des systèmes intentionnels, les systèmes fonctionnels sont aux croyances et aux désirs ce que les systèmes physiques sont aux centres de gravité et le principe de rationalité est aux croyances et désirs ce que l'attraction universelle est aux centres de gravité. En matière d'états et de propriétés fonctionnels, Dennett ne semble pas être instrumentaliste. S'il a pu dire qu'adopter le point de vue intentionnel vis-à-vis d'un jeu d'échecs revenait à le traiter *comme si* il avait des croyances et des

désirs, il n'a jamais soutenu qu'adopter le point de vue de la conception vis-à-vis du système revenait à le traiter *comme si* il avait des états fonctionnels: l'ordinateur à des états fonctionnels. Cependant, puisque Dennett refuse le fonctionnalisme turingien, son attitude réaliste vis-à-vis des états et propriétés fonctionnels ne suffit pas à faire de lui un réaliste, même modéré, en matière de désirs et croyances. Pour pouvoir se proclamer réaliste modéré vis-à-vis des croyances et désirs considérés comme *abstracta*, Dennett doit encore être réaliste vis-à-vis du principe de rationalité. Autrement dit, Dennett ne peut pas se contenter de considérer la rationalité comme une hypothèse normative dont a besoin l'interprète, il doit postuler que les systèmes interprétés sont effectivement rationnels. Autrement dit, pour pouvoir se dispenser du Réalisme en matière d'attitudes propositionnelles, il faut devenir Réaliste en matière de Rationalité.

En bref, en refusant un réalisme de type fodorien, c'est-à-dire en refusant une réalité ontologique aux attitudes propositionnelles, Dennett doit renoncer à leur faire jouer un rôle causal. Par ce refus, il se prive d'une explication causale possible du pouvoir prédictif des attributions d'attitudes propositionnelles. Il lui reste le choix entre deux options. Ou bien, il s'avoue instrumentaliste, et, dans ce cas, il peut certes engranger les bénéfices prédictifs que procure l'adoption du point de vue intentionnel mais doit renoncer à en donner une explication, ce qui peut passer pour une sérieuse entorse au code de déontologie philosophique. Ou bien, c'est la solution pour laquelle il semble opter dans *The Intentional Stance*, il défend un réalisme modéré. Mais nous venons de la voir, adopter un tel réalisme revient à expliquer le succès prédictif des attributions intentionnelles par la rationalité des systèmes interprétés. Le réaliste modéré doit donc faire la preuve de cette rationalité.

### **3. Evolution, rationalité et intentionnalité: le paradigme panglossien**

Selon Dennett, l'adoption de la perspective intentionnelle est justifiée à chaque fois que nous avons des raisons de supposer que l'hypothèse de la conception optimale est garantie. L'hypothèse de l'optimalité de la conception apparaît ainsi comme le relais qui permet de mettre en relation des états mentaux intentionnels considérés comme des *abstracta* ayant un rôle purement prédictif et des états fonctionnels, non-intentionnels, de l'organisme dont le rôle est causal. L'hypothèse de l'optimalité de la conception trouve aux yeux de Dennett sa garantie dans la théorie darwinienne de l'évolution. La théorie de l'évolution présente un second avantage: elle "montre comment chaque trait du monde naturel peut être le produit d'un processus aveugle, imprévoyant, non téléologique, et en définitive mécanique de reproduction différentielle à travers de longues périodes de temps" (Dennett,

1987, trad. fr.: 366-67). Cette seconde caractéristique sera à la base de l'argument de Dennett contre la doctrine de l'intentionnalité intrinsèque. Je n'examinerai pas ici directement cet argument. J'examinerai en revanche certains des postulats sur lesquels il repose et vérifierai le bien-fondé des affirmations de Dennett quant au lien existant entre évolution et rationalité.

### **Evolution et rationalité: première esquisse**

Ainsi que l'a fait remarquer Stich (1990), la stratégie de Dennett et plus généralement les stratégies qui font appel à la théorie de l'évolution pour défendre la thèse selon laquelle les systèmes cognitifs humains sont en grande partie rationnels, reposent sur deux idées. La première est que l'évolution produit des systèmes dont la conception est optimale ou proche de l'optimal. La seconde est qu'un système cognitif optimalement conçu est un système cognitif rationnel.

On peut reconstituer l'argument qui mène à la première de ces idées de la manière suivante. D'un point de vue biologique, un système est mieux conçu qu'un second, si les individus qui possèdent le premier système ont de meilleures chances de survie et de reproduction que les individus qui possèdent le second. Un système optimalement conçu est donc un système qui assure à ses possesseurs de meilleures chances de survie et de reproduction que tout autre système possible. Pourquoi devrions nous penser que l'évolution produira des systèmes optimalement conçus ou proches de l'optimalité? La réponse se fonde sur deux thèses: (1) la sélection naturelle est le moteur de l'évolution et (2) le hasard des mutations génétiques fournit à long terme, un éventail très ouvert et diversifié d'options. Si la sélection naturelle est le moteur de l'évolution, on peut compter sur elle pour se comporter comme un agent optimisateur. Autrement dit, si deux espèces sont en concurrence ou si, au sein d'une espèce, deux phénotypes sont en concurrence, on peut être sûr que c'est le système le mieux conçu qui sera sélectionné. En outre, le caractère aléatoire des variations génétiques et, sur une longue durée, le très grand nombre de ces mutations, garantissent que des gènes mutants codant des conceptions optimales ou proches de l'optimal verront le jour. Au cours du temps, la sélection naturelle aura donc à choisir entre de nombreux systèmes dont certains seront proches de l'optimalité. Il est par conséquent vraisemblable que le système auquel l'évolution aura abouti au terme d'une longue période de temps sera optimal ou proche de l'optimal. Pour que ceci s'applique aux systèmes cognitifs, il suffit d'admettre que les systèmes cognitifs sont eux aussi le produit de l'évolution.

De la seconde idée, l'idée selon laquelle un système cognitif optimalement conçu est un système cognitif rationnel, Dennett n'a pas donné de justification

explicite. Cependant, à partir de passages comme celui-ci, "la capacité d'avoir des croyances n'aurait pas de valeur de survie s'il ne s'agissait d'une capacité à avoir des croyances vraies" (Dennett, 1978: 17), on peut être tenté de reconstituer l'argument suivant: (1) un système rationnel est un système qui produit des croyances vraies; (2) le fait d'avoir des croyances vraies augmente les chances de survie et de reproduction; (3) les systèmes cognitifs choisis par la sélection naturelle seront des systèmes rationnels.

Sous cette forme abrupte, l'argument prête le flanc à la critique. Stich (1990) formule des objections convaincantes contre cette version de l'argument. L'aptitude à la survie d'un système ne se juge pas à sa seule capacité à produire des croyances vraies, des représentations correctes de l'état de son environnement; elle se juge également à l'économie de moyens avec laquelle il y parvient. Or, il n'est pas certain que les impératifs de la survie soient toujours compatibles avec le loisir et les moyens qu'exige la recherche de la vérité. La gazelle qui s'enfuirait au premier bruit suspect aurait sans doute moins de croyances vraies que sa consœur qui vérifie avant de s'enfuir que ce bruit traduit bien la présence d'un danger, mais la seconde risquerait de faire les frais de sa rationalité supérieure en servant de dîner au lion. En bref, la vérité a un prix, et passé un certain seuil, elle risque de nous coûter trop cher.

Il est bien sûr possible de neutraliser ces objections en définissant la rationalité non pas en termes de capacité absolue à produire des croyances vraies mais en termes de capacité à trouver le meilleur équilibre possible entre bénéfices (croyances vraies) et coûts (dépenses d'énergie et de temps). Il devient alors trivialement vrai qu'un système optimalement conçu est un système optimalement rationnel, puisque définition de la rationalité optimale et définition de la conception optimale coïncident. La conception a pour critère l'aptitude à la survie, l'aptitude à la survie a pour critère le rapport coût/bénéfice dans le comportement du système et le rapport coût/bénéfice est à son tour le critère de la rationalité. Sera donc à la fois optimalement conçu et optimalement rationnel le système qui aura trouvé le point d'équilibre entre coûts et bénéfices cognitifs. Mais ce nouvel argument contient une prémisse implicite, à savoir qu'il existe entre coûts et bénéfices un unique point d'équilibre et non une pluralité de points d'équilibre. Si plusieurs équilibres sont possibles, on ne peut parler ni de LA conception optimale ni de LA rationalité optimale. On doit parler des conceptions optimales et des rationalités optimales.

Sur le projet de Dennett - faire de la sélection darwinienne la théorie relais entre une ontologie des systèmes cognitifs qui n'admet pas en son sein les attitudes propositionnelles et une théorie de l'interprétation qui exploite les succès prédictifs de l'attribution d'attitudes intentionnelles - plane une double menace. La première est liée à la remise en cause par les nouvelles théories de l'évolution de

l'adaptationnisme classique et du primat de la sélection naturelle comme moteur de l'évolution. Il semble que si la sélection naturelle n'est qu'un facteur parmi d'autres de l'évolution, on ne puisse plus présupposer que l'évolution produise nécessairement des systèmes optimalement conçus. La deuxième menace est celle de la diversité cognitive. S'il n'est pas possible d'écarter l'idée que la rationalité n'est pas simplement une affaire de degré, si l'on doit admettre que des systèmes cognitifs différents puissent manifester des types différents de rationalité, il semble que le succès prédictif des attributions intentionnelles ne puisse plus s'expliquer simplement par la coïncidence entre le postulat de rationalité qui guide l'interprétation et la rationalité effectivement manifestée par les systèmes interprétés. Si Dennett n'a apparemment pas pris en compte la seconde menace, en revanche il n'a pas ignoré la première. Pour la contrer, il a défendu un adaptationnisme modéré, compatible selon lui avec les données dont les adversaires de l'adaptationnisme se sont servis dans leurs arguments. Je présenterai tout d'abord la version de l'adaptationnisme défendue par Dennett et montrerai que même si on l'accepte, le projet de Dennett succombe à la menace de la "diversité cognitive". Je montrerai ensuite que la reconstruction adaptationniste de Dennett constitue une distorsion des données et ne peut être acceptée.

### **L'adaptationnisme de Dennett**

L'argument en faveur de l'optimalité de la conception que nous avons esquissé plus haut se fondait sur les thèses de l'adaptationnisme classique. Selon cette conception classique, l'évolution est la résultante de deux facteurs: un facteur-amont et un facteur-aval. Par facteurs-amont, j'entends les facteurs responsables de la venue à l'existence de formes nouvelles, autrement dit, les facteur-amonts sont les pourvoyeurs d'options. Dans l'adaptationnisme classique, le facteur-amont est unique: il s'agit du hasard des mutations génétiques. Par facteurs-aval, j'entends les facteurs responsables des choix effectués parmi ces options, ou plutôt, car le terme de choix n'est pas neutre, les facteurs qui expliquent la persistance de certaines options et la disparition des autres. L'adaptationnisme classique n'admet qu'un facteur-aval: la sélection naturelle.

Cette version adaptationniste de la théorie de l'évolution fait, depuis un certain nombre d'années, l'objet de critiques sévères. Ainsi, Gould et Lewontin (1979) s'élèvent-ils contre le panglossianisme auquel conduit l'adaptationnisme. Comme le docteur Pangloss de Voltaire, les adaptationnistes voient en ce monde le meilleur des mondes possibles et en donnent pour raison le rôle moteur de la sélection naturelle: si un trait existe, c'est qu'il a été sélectionné et s'il a été

sélectionné, c'est qu'il avait une valeur adaptative. Contre cet optimisme adaptationniste, Gould et Lewontin font valoir que la sélection naturelle n'est qu'un facteur parmi d'autres de l'évolution et que sont à prendre en compte les conditions historiques et architectoniques de l'évolution, par exemple, des phénomènes tels que la dérive génétique, l'inertie phylétique, la compensation matérielle, les contraintes morphologiques et embryologiques. Ainsi, au cours de l'évolution d'une espèce, le nombre de doigts que ses membres possèdent peut varier, augmenter ou diminuer, mais l'ordre d'apparition ou de disparition n'est pas aléatoire, il est fixé par des contraintes architectoniques. De la même manière l'ordre d'apparition des variations peut influencer sur le résultat final de l'évolution. Soit deux traits A et B. Supposons que A ait une valeur adaptative propre mais que B ne contribue à rendre l'organisme plus apte que s'il est conjugué à A. Si la mutation codant B apparaît après que A ait été sélectionné, B sera également sélectionné, mais si la mutation codant B apparaît la première, B ne sera pas sélectionné. A un adaptationnisme pur et dur, Gould et Lewontin proposent donc de substituer une vision pluraliste de l'évolution qui admet des explications non-adaptationnistes de l'existence de certains traits.

Dennett s'efforce de minimiser la portée de ces objections et pour ce faire adopte une double stratégie: la première consiste à défendre l'idée que le rôle des hypothèses d'optimalité est avant tout d'ordre méthodologique et heuristique; la seconde, liée à la première, consiste à suggérer que l'adaptationnisme reste compatible avec l'idée que "Dame Nature n'est pas un agent d'optimisation mais de satisfaction" (Dennett, 1987, trad. fr.: 341).

Le principal reproche adressé par Gould et Lewontin au programme adaptationnisme est d'être infalsifiable. Ainsi qu'Amundson (1990: 577) le fait remarquer, le terme d'adaptationnisme renvoie à la fois à un programme de recherche sur l'évolution et à la croyance factuelle centrale sur laquelle se fonde ce programme. La croyance factuelle est que la plupart des caractères des organismes constituent des adaptations à l'environnement et ont été sélectionnés pour cette raison. Le programme utilise cette croyance pour guider la recherche: on suppose qu'une structure donnée constitue une adaptation et l'on s'efforce d'en donner une explication sélective compatible les données connues de l'histoire naturelle.

Dans un premier temps Dennett s'efforce de défendre l'adaptationnisme en tant que programme de recherche: en tant que programme de recherche l'adaptationnisme est effectivement infalsifiable, mais cette infalsifiabilité n'a rien que de bénin. Il souligne que le problème commun aux psychologues et aux biologistes est le problème de l'interprétation de leurs données respectives. Ce travail d'interprétation ne saurait être mené à bien sans qu'on fasse appel à des postulats d'optimalité: postulat de rationalité pour le psychologue, postulat adaptationniste dans le cas du



biologiste: "Nous adoptons les postulats d'optimalité non pas parce que nous pensons naïvement que l'évolution a fait de ce monde le meilleur des mondes possibles, mais parce que nous devons être des interprètes, si nous voulons faire un progrès quelconque, et l'interprétation requiert que l'on invoque l'optimalité" (Dennett, 1987, trad. fr.: 360). Une dialectique subtile s'instaure entre biologie et psychologie. D'une part, le postulat d'optimalité de la pensée adaptationniste apparaît comme "une version spéciale du point de vue intentionnel en théorie de l'évolution" (Dennett, 1987, trad. fr.: 366). Pour interpréter les phénomènes naturels existants, on postule que la Nature est un agent rationnel. D'autre part le principe de rationalité, qui commande l'attribution à autrui d'états mentaux, apparaît à son tour comme un cas particulier du principe d'optimalité en biologie: les systèmes cognitifs humains sont considérés comme un cas particulier de systèmes naturels optimalement conçus.

En insistant sur le rôle méthodologique de l'hypothèse d'optimalité et interprétant les critiques de Gould et Lewinson comme des critiques visant l'adaptationnisme en tant que programme de recherche, Dennett redessine à sa manière la ligne de partage entre théories adaptationnistes de l'évolution et théories "structuralistes" de l'évolution. Selon lui, la stratégie adaptationniste pose la question aristotélicienne "pourquoi?" et cherche en réponse des raisons, tandis que la stratégie structuraliste ignore le pourquoi pour s'intéresser au comment et recherche les causes de ce qui s'est passé. Dennett peut ensuite dénoncer l'illusion que le rejet de la recherche des raisons au profit de la recherche des causes soit un progrès et arguer du fait qu'une enquête sur les causes de ce qui s'est passé demande qu'on ait au préalable catégorisé ce qui s'est passé, catégorisation qui est fonctionnelle et qui se fonde nécessairement sur des réponses aux questions "Pourquoi?"

Dennett souligne que toute recherche doit être guidée par une interprétation et que la question "Pourquoi?" constitue un guide précieux pour la recherche des causes. Une explication causale d'un comportement ou d'un trait biologique ne requiert pas la fausseté d'une explication par des raisons. Une telle vue n'a rien de révolutionnaire en épistémologie: elle n'est qu'une version particulière de l'idée selon laquelle la recherche ne procède pas simplement par théorisation directe sur les données brutes d'observation mais requiert que l'on fasse des hypothèses qui vont guider l'observation et nous permettre de nous orienter parmi ces données. Cependant, Dennett a tort d'imputer aux théories structurelles de l'évolution le rejet de toute explication par les raisons: ces théories condamnent l'abus de telles explications, non leur usage. Elles ne remettent pas en cause l'utilité heuristique de la question "Pourquoi?" mais l'idée qu'à cette question on peut toujours répondre en donnant une bonne raison.

Gould et Lewontin ne s'en prennent pas à l'infalsifiabilité de l'adaptationnisme en tant que programme de recherche, mais à son infalsifiabilité en tant que croyance centrale. Les adaptationnistes se défendent de ce travers en soulignant que chaque explication adaptationniste particulière est falsifiable. Mais ce n'est pas là le fond du reproche qui est fait à l'adaptationnisme. Les critiques de l'adaptationnisme demandent que l'on considère des types concurrents d'explication et pas seulement des explications concurrentes appartenant à un même type.

Par conséquent, même si certaines des critiques adressées par Gould et Lewontin au programme adaptationniste peuvent être jugées excessives, il est tout aussi excessif de placer - comme le fait Dennett - Gould et Lewontin dans le camp de Skinner en assimilant leur attitude vis-à-vis de l'adaptationnisme à celle de Skinner vis-à-vis du mentalisme. Contrairement à ce qu'affirme Dennett, Skinner ne reprochait pas tant au mentalisme de faire appel à des explications par des raisons que de donner des explications faussement causales en invoquant de prétendues entités inobservables (croyances, désirs et intentions). Bien au contraire, Gould et Lewontin reprochent à l'adaptationnisme le fait qu'en cherchant toujours des raisons, il attache trop d'importance à la sélection naturelle comme cause de l'évolution et néglige les autres causes possibles.

Le fait que les hypothèses d'optimalité soient indispensables pour la compréhension des phénomènes biologiques et psychologiques ne suffit pas à démontrer que les systèmes biologiques existants, en particulier cognitifs, sont *dans les faits* optimalement conçus. La nécessité de l'hypothèse d'optimalité comme principe méthodologique est compatible avec sa fausseté comme loi de la Nature. Or Dennett ne peut se contenter de défendre l'adaptationnisme d'un point de vue méthodologique: il a précisément besoin que cette hypothèse soit confirmée dans les faits si la théorie de l'évolution doit servir de relais et venir expliquer les succès prédictifs liés à l'adoption de la perspective intentionnelle. Dennett doit donc défendre non seulement un adaptationnisme méthodologique mais un adaptationnisme factuel. A cette vision panglossienne de l'adaptationnisme, les critiques venues de l'école structuraliste opposent le fait que la plupart du temps les solutions adoptées par la nature ne sont pas celles que l'ingénieur jugerait optimales mais constituent, selon l'expression de François Jacob (1977), une forme de "bricolage": la nature fait ce qu'elle peut avec les moyens du bord.

A ces objections, Dennett répond en soulignant qu'entre l'optimalité et les différentes contraintes prises en compte un marchandage est toujours possible: "Ce qui peut paraître comme loin de l'optimum d'après un certain ensemble de conditions peut sembler optimal d'après un ensemble plus large de conditions". (Dennett, 1987, trad. fr.: 340). Dennett prend ainsi la défense d'un adaptationnisme modéré: la

nature n'est pas parfaite mais elle fait ce qu'elle peut. Etant donné les contraintes structurelles et historiques, le nombre d'options entre lesquelles choisir est limité, mais parmi les options disponibles, la sélection naturelle choisit la meilleure. Autrement dit, la stratégie de Dennett repose sur un choix de factorisation: toutes les contraintes structurelles ou historiques de l'évolution mises en évidence par les non-adaptationnistes sont ramenées au rang de facteurs-amont, la sélection naturelle demeure le seul facteur-aval. Ainsi que nous le verrons, il est loin d'être évident que l'adaptationnisme modéré prôné par Dennett soit la bonne manière de conceptualiser et pour ainsi dire de neutraliser les données biologiques semblant confirmer l'existence d'une évolution non adaptative. Cependant, même si l'on néglige pour l'instant la possibilité que la théorisation proposée par Dennett de ces données soit erronée, l'adaptationnisme n'est pas sans conséquence pour le point de vue intentionnel.

Si les systèmes biologiques ne sont optimaux que relativement à un arrière-plan de contraintes, il ne suffit plus pour interpréter les données biologiques d'adopter le principe d'optimalité, il faut encore connaître les contraintes auxquelles sont soumises les configurations biologiques examinées. Mais quelle connaissance le naturaliste a-t-il de ces contraintes? Comment les identifie-t-il? Quand devra-t-il considérer une configuration comme non-optimale? Quand devra-t-il réviser sa liste de contraintes? On peut craindre ici de fâcheux glissements. Dennett (1987, 1990: 344) soutient que "nous réussissons en réalité très bien à choisir les bonnes contraintes" et que l'on peut prouver que l'on a localisé "toutes les contraintes authentiquement importantes en arguant du fait que:

- (1) la configuration optimale étant donné ces contraintes est A
- (2) Dame Nature optimise
- (3) A est la configuration observée"

Un tel argument ressemble dangereusement à une pétition de principe. Dennett reconnaît cette fâcheuse ressemblance mais distingue deux points de vue. Selon un premier point de vue, il y a effectivement circularité puisque le caractère optimisateur de la nature est présupposé: il n'apparaît pas comme la conclusion de l'argument mais comme une de ses prémisses. Mais d'un autre point de vue, il n'y a pas circularité vicieuse, car le caractère optimisateur de la nature est accepté en tant que postulat méthodologique et non en tant que thèse factuelle. Il y aurait effectivement circularité vicieuse, si les listes de contraintes étaient toujours révisées *post hoc* pour faire coïncider hypothèse d'optimalité et réalité factuelle. Si les adaptationnistes s'autorisaient une telle latitude, l'adaptationnisme serait une imposture. Mais précisément, nous dit Dennett, ils ne le font pas. Que se passe-t-il alors lorsque la configuration observée ne correspond pas à la configuration

optimale calculée en prenant compte des contraintes et que l'on se refuse pourtant à réviser la liste de ces contraintes? Dans un tel cas, on doit conclure, selon Dennett, à une indétermination fonctionnelle de la configuration observée et reconnaître qu'au niveau des faits, le principe d'optimalité est à strictement parler faux. N'est-ce pas là un aveu d'échec? La stratégie de Dennett se retourne-t-elle pas contre lui?

Bien au contraire, répond Dennett, l'existence de cas d'indétermination fonctionnelle ne fait que renforcer le parallèle entre biologie et psychologie. Rappelons que Dennett considère l'attribution de croyances et désirs comme une entreprise d'interprétation radicale et voit dans l'indétermination des contenus un argument contre le réalisme. Si l'utilisation du principe de rationalité pour l'interprétation des données observables ne permet pas d'éliminer toutes les interprétations sauf une, alors plus rien ne peut venir départager les interprétations restantes. La mise en évidence de cas d'indétermination fonctionnelle ne peut que conforter Dennett dans son adhésion à la thèse d'indétermination quinienne. Pour pouvoir attribuer des contenus d'attitudes propositionnelles, il faut que nous soyons capables d'assigner une fonction particulière à certains états de l'organisme. Si l'assignation des fonctions peut être sous-déterminée par les contraintes existantes, il en ira de même de l'attribution de contenus. En bref, l'indétermination fonctionnelle sous-tend l'indétermination des contenus et la thèse d'indétermination de Quine est biologiquement fondée. Mais cette apparente complaisance de la biologie à l'égard des thèses de Dennett n'est-elle pas un effet d'optique?

### **Le dogme d'unicité**

L'adaptationnisme avoué de Dennett s'accompagne d'une adhésion inavouée à un dogme d'unicité. Les deux thèses ne sont en rien indissociables. Certaines des conclusions auxquelles Dennett aboutit, en particulier l'idée d'indétermination fonctionnelle, semblent découler du dogme d'unicité bien plus directement que de l'adaptationnisme. Or, si l'on peut avancer des arguments biologiques en faveur de l'adaptationnisme, tel n'est pas le cas en ce qui concerne le dogme unitaire.

Lorsque rationalité et optimalité sont prises comme des principes normatifs absolus, il est parfaitement légitime de postuler l'unicité de la norme. Des désaccords peuvent surgir quant à la nature de la norme, ou quant à la possibilité même de la définir, mais on ne peut guère contester qu'en tant qu'absolu ou idéal, l'idée de norme comporte l'idée d'unicité. Si l'on utilise les termes de rationalité ou optimalité non plus pour désigner la norme elle-même mais pour qualifier un type d'objet possible ou réel (système, configuration) en tant qu'il est conforme à la norme, il demeurerait légitime de postuler l'unicité, s'il était possible qu'un type d'objet se conforme à la

norme sans restriction aucune, autrement dit si le passage de l'idéal au réel pouvait s'effectuer sans solution de continuité. Néanmoins, si le réel est disjoint de l'idéal, s'il impose ses contraintes propres aux objets, la conformité à la norme ne peut être atteinte de manière absolue. On ne peut alors conclure de l'unicité de la norme à l'unicité du type d'objet susceptible de l'approcher au plus près. Dans un tel cas, l'affirmation selon laquelle l'optimalité factuelle peut être définie de manière unique constitue une affirmation dogmatique, c'est ce que j'appelle le dogme d'unicité.

Lorsque Dennett fait appel à l'optimalité et à la rationalité comme principes méthodologiques, il les utilise comme normes et rien ne s'oppose à ce que, en tant que normes, rationalité et optimalité aient un caractère absolu. En revanche, comme Dennett est prompt à le reconnaître, ces normes ne sont pas parfaitement réalisées dans la réalité. La rationalité telle qu'elle se manifeste factuellement est une rationalité limitée et l'optimalité est une optimalité sous contraintes. Pourtant, Dennett fait sien le dogme d'unicité lorsqu'il affirme que nous pouvons calculer ce que serait *la* configuration optimale ou *la* croyance rationnelle étant donné les contraintes. Ce faisant, il traite les contraintes comme un phénomène linéaire: les contraintes s'additionnent, chacune apportant une restriction supplémentaire, mais elles n'interagissent pas entre elles.

Cette vision des choses conduit Dennett à largement sous-estimer les difficultés que comporte un calcul d'optimalité en biologie. Il ne suffit pas pour effectuer ce calcul de trouver un juste compromis entre d'un côté les différentes contraintes et de l'autre l'optimalité. Il est également nécessaire de trouver des compromis entre différentes contraintes et l'on ne peut estimer *a priori* que de tous les compromis possibles, un seul serait le compromis idéal. Le cas de certains poissons d'eau-vive de la famille des cyprinidés, que rapportent Kitcher et Kitcher (1988), fournit une illustration intéressante de la complexité de ces contraintes. Les poissons mâles de l'espèce ont les flancs tachetés. La taille et la couleur des taches peut varier considérablement d'une population à une autre et à l'intérieur même d'une population. Les études menées par John Endler (1983) ont montré que les femelles semblent préférer les mâles qui ont les taches les plus grandes et les plus brillantes. Est-ce à dire que pour les poissons mâles, la configuration optimale consisterait à avoir les tâches les plus grandes et les plus brillantes possibles? Non point, car si les taches sont trop grandes ou trop brillantes, les poissons sont plus facilement repérés par leurs prédateurs. Le calcul d'optimalité doit donc prendre en compte cette contrainte: les taches doivent être grandes et brillantes mais pas trop. La prise en compte de cette contrainte, permet-elle de calculer une configuration optimale, disons, des taches d'un diamètre de sept millimètres? Que dire des mâles dont les taches font six millimètres de diamètre et de ceux dont les taches sont de huit

millimètres de diamètre? Les premiers sont un peu moins attirants pour les femelles mais ils sont moins vulnérables aux prédateurs, les seconds sont plus vulnérables aux prédateurs mais ont également plus de succès auprès des femelles. Sans doute, les poissons dont les taches seraient vraiment trop remarquables seraient-ils dévorés avant d'avoir pu se reproduire, tandis que les poissons trop ternes finiraient sans descendance une longue vie paisible. Mais entre ces deux extrêmes on trouve non pas une configuration optimale mais une série de configurations également proches de l'optimal. Encore ai-je ici simplifié l'exemple à outrance. En fait, toutes une série d'autres contraintes entrent en jeu: la coloration du lit du cours d'eau, les conditions d'éclairage, les types de prédateurs rencontrés, l'acuité visuelle de ces prédateurs. En outre, on doit ajouter à ces contraintes écologiques des contraintes internes aux organismes. Ainsi, les variations possibles de la forme des taches semblent-elles aussi soumises à des contraintes développementales. Plus le nombre de contraintes augmente, plus s'accroissent le nombre et la variété des solutions susceptibles d'avoir une valeur adaptative sensiblement égale. Plus généralement, le simple fait qu'un très grand nombre de populations apparemment stables manifestent une grande diversité tant au niveau des génotypes qu'à celui des phénotypes devrait inciter à penser que même en matière d'optimalité, l'unicité est l'exception plutôt que la règle.

En quoi ces conclusions remettent-elles en cause les thèses d'indétermination soutenues par Dennett? Commençons par l'indétermination fonctionnelle. Selon Dennett, si la configuration observée ne correspond pas à la configuration calculée d'après une liste de contraintes jugée exhaustive, nous nous trouvons devant un cas d'indétermination fonctionnelle. Mais, si plusieurs configurations sont également adaptatives, le fait que la configuration observée ne corresponde pas à la configuration optimale calculée, ne prouve pas que cette configuration ne soit pas optimale. Il peut aussi bien indiquer que le calculateur n'a pas su correctement apprécier toute la subtilité du jeu des contraintes et n'a énoncé qu'une des solutions possibles du problème. En outre, comme l'a montré l'exemple des cyprinidés, les contraintes peuvent être très spécifiques et subir des variations importantes d'un cas à un autre. Ici, Dennett semble pris à son propre piège: s'il reconnaît cette spécificité, sa défense de l'adaptationnisme s'écroule. Nous avons vu en effet que sa défense de l'adaptationnisme met en avant le rôle prédictif de l'hypothèse d'optimalité. Puisqu'il présuppose l'optimalité au lieu d'essayer d'en donner une démonstration, la seule parade dont il dispose face aux accusations de vacuité est d'insister sur le caractère stable de la liste des contraintes à partir de laquelle s'effectuent les calculs optimisateurs. Sa thèse d'indétermination fonctionnelle apparaît donc à la fois comme un artefact de la manière dont il a choisi de défendre l'adaptationnisme et

comme un effet du dogme d'unicité qui le rend aveugle à la multiplicité des voies que la nature peut choisir d'emprunter.

Qu'en est-il alors de la thèse d'indétermination des contenus? Si l'optimalité que l'on peut attendre des configurations produites par l'évolution est une optimalité sous contraintes et si les systèmes cognitifs sont eux-mêmes les produits de l'évolution, la rationalité qu'il manifesteront sera une rationalité limitée. Si les systèmes cognitifs ont une rationalité limitée, il ne suffit plus d'adopter le principe de rationalité pour prédire ou expliquer les comportements. Il faut encore connaître les contraintes auxquelles est soumis l'organisme auquel on prête des états intentionnels. Le cas de l'attributeur est ici parallèle à celui du naturaliste. On doit se demander quelle connaissance il a des contraintes, selon quels principes il décide s'il doit en réviser la liste ou considérer un comportement comme irrationnel.

Selon Dennett, si l'utilisation du principe de rationalité pour l'interprétation des données observables ne permet pas d'éliminer toutes les interprétations sauf une, nous nous trouvons devant un cas d'indétermination radicale. On peut penser que les contraintes écologiques, externes, font partie des données observables. Mais la biologie de l'évolution, nous apprend qu'à côté de ces contraintes externes, existent des contraintes structurales internes. Si le projet de Dennett est de donner un fondement darwinien à sa théorie des systèmes intentionnels, il doit accepter les données que fournit la biologie de l'évolution et admettre l'existence de contraintes structurales. Mais s'il admet l'existence de contraintes structurales, Dennett ouvre la voie à une autre analyse: il est possible que l'indétermination qu'un calcul rationnel fondé sur les seules contraintes externes ne permet pas d'éliminer soit levée si sont prises en compte les contraintes structurales.

On peut ici utiliser une stratégie similaire à la stratégie mise en oeuvre par Chomsky (1969) dans sa réponse à l'argument quinien de l'indétermination de la traduction. Chomsky soulignait que si, effectivement, la seule induction à partir des comportements verbaux observables ne permettait pas de départager tous les systèmes rivaux d'hypothèses, elle ne permettait pas non plus d'expliquer l'apprentissage du langage. Il montrait que pour en rendre compte, il fallait supposer que seules certaines de ces hypothèses étaient accessibles à l'enfant, autrement dit supposer que certaines contraintes structurales déterminaient le type d'hypothèses susceptibles d'être mises en oeuvre.

La même question se pose au sujet de l'apprentissage des attributions intentionnelles. Comment expliquer que, comme le confirment certains résultats expérimentaux en psychologie (Gelman *et al.* 1983; Hirschfeld, 1988; Turiel et Davidson, 1986), les enfants se comportent comme des attributeurs, qu'ils intériorisent très rapidement et sans instruction explicite les principes de la

psychologie naïve? S'il n'est pas possible d'attribuer des états intentionnels au moyen simplement de l'hypothèse de rationalité et de l'induction à partir des données observables, si en outre il est nécessaire de tenir compte d'un certain nombre de contraintes internes, comment l'enfant apprend-il à localiser si vite les contraintes importantes?

Une des réponses possibles consisterait à substituer au principe de rationalité le principe d'empathie. Certains philosophes<sup>2</sup> ont soutenu que la différence entre les deux principes était purement nominale, s'agissant dans un cas comme dans l'autre de la projection sur autrui de rationalité. Des différences persistent cependant: ce qui selon le principe de rationalité est projeté sur autrui est une Rationalité "majuscule"; ce qui selon le principe d'empathie est projeté est le type de rationalité contrainte propre au système cognitif de l'attributeur. S'il veut prédire avec succès le comportement d'un système dont la rationalité est limitée, l'attributeur qui fait usage du principe de rationalité doit préalablement avoir une connaissance explicite des limites du système considéré et des contraintes auxquelles il obéit. En revanche, si l'attributeur fait usage du principe d'empathie, ces connaissances préalables ne sont pas requises aussi longtemps que le système considéré est assez semblable au système de l'attributeur.

Si cette hypothèse est juste, le fait que dans bien des cas nous puissions par l'attribution de croyances et désirs expliquer et prédire correctement le comportement d'autrui n'est pas la preuve que le système interprété soit optimalement conçu et donc rationnel; il indique simplement que le type de rationalité contrainte propre au système de l'attributeur est suffisamment semblable à celui du système interprété. De la même manière le fait que nous puissions dans certains cas attribuer à autrui des contenus mentaux déterminés ne prouve pas que nous nous soyons trouvé dans le cas favorable où une unique interprétation rationnelle était possible. Peut-être plusieurs interprétations rationnelles étaient-elles possibles, une seule étant toutefois accessible à l'interprète étant donné le type de contraintes pesant sur sa rationalité. Enfin, inversement, le fait qu'un interprète ne

---

<sup>2</sup> La généalogie des deux principes est assez déroutante. Quine peut être considéré comme le père de l'un et de l'autre. On les trouve exposés à quelques pages d'intervalle dans la section 45 de *Le Mot et la Chose*. Pour compliquer les choses, Quine attribue à des conversations avec Davidson - pourtant connu comme défenseur du principe de rationalité - l'origine de l'idée du principe d'empathie. Levin (1989) soutient que les versions modérées des deux principes tendent à se fondre en une seule et même conception. Dans son article "Making Sense of Ourselves" (repris dans Dennett, 1987), Dennett soutient plus radicalement que la différence entre les deux principes est tout au plus une différence d'accent.



parvienne pas à attribuer un contenu déterminé aux croyances et désirs du système interprété ne prouve pas que ce contenu soit effectivement indéterminé. Si cette incapacité à attribuer des contenus déterminés à autrui est due à une différence du type de limites pesant sur la rationalité du système interprété et sur celle du système interprétant, l'indétermination n'est pas forcément dans le système interprété, elle peut être dans l'oeil de l'interprète. L'interprète n'est pas d'essence divine: nous sommes tous des interprètes et tous limités dans notre rationalité.

A cela Dennett pourrait encore répondre que ces objections ne valent pas si l'on s'en tient à l'interprétation telle qu'elle se pratique au sein de l'espèce humaine. Son argument pourrait prendre la forme suivante. Si les contraintes externes et donc observables peuvent varier, tous les membres de l'espèce humaine sont en revanche soumis aux mêmes contraintes internes. Par conséquent, si la rationalité des systèmes cognitifs humains est limitée du fait de ces contraintes internes, les limites en question sont les mêmes pour tous les systèmes. Lorsque nous interprétons le comportement d'autres espèces animales et ne parvenons pas à attribuer à ces organismes des contenus intentionnels déterminés, nous pouvons faire valoir, pour rendre compte de cette indétermination apparente, notre ignorance des contraintes internes auxquelles ils sont soumis. En revanche, une telle explication n'est pas possible dans le cas humain. Si l'indétermination des contenus intentionnels attribués à d'autres êtres humains n'est pas réductible à une différence de contraintes internes, cette indétermination reste radicale. Dennett n'a alors aucune raison d'abandonner ses positions anti-réalistes en matière de croyances et désirs.

Cette réponse hypothétique de Dennett présuppose le dogme d'unicité. Si la diversité génétique est la règle et si, comme le veut Dennett, nos capacités cognitives sont génétiquement déterminées, on ne peut pas poser sans donner d'arguments que les systèmes cognitifs font exception à la règle. Même en supposant que les systèmes cognitifs humains sont innés, même en supposant qu'au cours de l'évolution ils sont tous été soumis aux-mêmes contraintes et qu'ils sont tous optimaux relativement à ces contraintes, il ne s'ensuit pas qu'ils sont identiques. Comme l'a montré l'exemple des cyprinidés, il est possible que des stratégies génétiquement codées très diverses aient été développées face à ces contraintes et que plusieurs de ces stratégies puissent prétendre à l'optimalité.

Nous avons montré jusqu'ici que, même si l'on accorde un rôle privilégié au principe d'optimalité, l'adaptationnisme ne fournit pas de soubassement biologique au principe d'indétermination quinien. Les arguments de Dennett en faveur d'une indétermination fonctionnelle présupposent en outre l'unicité de la configuration optimale. Or la biologie de l'évolution nous apprend non certes qu'il ne peut jamais y avoir de configuration optimale unique, mais qu'un tel cas de figure constitue

l'exception plutôt que la règle. L'évolution ne garantit donc pas l'unicité. Reste à se demander si elle garantit l'optimalité.

### **Les habits neufs de l'harmonie préétablie**

L'adaptationnisme classique explique les traits existants comme les résultats de variations engendrées au hasard et soumises à la sélection naturelle. Autrement dit, il ramène l'évolution à deux facteurs: en amont, le hasard qui à travers la variation fournit les options possibles et, en aval, la nécessité qui agissant sous la forme de la sélection naturelle choisit - rationnellement, dirait Dennett - entre les options proposées. Le renouvellement des théories de l'évolution a mis en évidence l'action d'autres forces évolutives. L'accent a été mis sur les conditions morphologiques et embryologiques de l'évolution, sur le caractère cumulatif de l'histoire évolutionniste - ce qui est possible à un moment ne l'est plus ensuite et la nature ne peut pas reculer pour mieux sauter -, sur des phénomènes tels que la dérive génétique, les mutations neutres, sur le rôle des migrations et de l'isolement reproductifs. Enfin, des biologistes comme Mayr ont insisté sur le caractère holiste et non atomiste de la sélection naturelle (c'est l'individu comme tout et non le gène qui est la cible de la sélection). Ces éléments nouveaux ne remettent pas en cause le fait que la sélection adaptative joue un rôle dans l'évolution mais elles contribuent à en relativiser la toute-puissance. La prise en compte de cet ensemble de facteurs conduit à reconsidérer les mécanismes de l'évolution. Cette réorganisation théorique peut prendre plusieurs formes.

L'approche défendue par Dennett consiste, nous l'avons vu, à considérer ces facteurs comme des éléments qui introduisent de la nécessité dans le hasard sans pour autant introduire de hasard dans la nécessité. Autrement dit ces facteurs sont analysés par Dennett comme des facteurs-amont: ils viennent contraindre le nombre et la forme des options possibles. En revanche, en ce qui concerne le choix entre ces options, la sélection naturelle reste seule maîtresse à bord. Le choix de cette approche n'est évidemment pas dénué d'arrière-pensées philosophiques. Ce n'est qu'en isolant dans le mécanisme évolutif un moment régi par la seule nécessité - le moment du choix - que Dennett peut préserver l'hypothèse de la conception optimale et avec elle celle de la rationalité, serait-ce au prix de quelques bémols, c'est-à-dire en concédant que le système n'est optimal que considéré globalement et non dans ses détails et que cette optimalité est relative à des contraintes.

L'approche de Dennett n'est justifiée que si l'on peut montrer, pour chacun des facteurs connus, qu'il agit uniquement en amont, qu'il intervient dans l'explication des contraintes qui pèsent sur les options susceptibles d'être offertes

mais ne joue aucun rôle dans l'explication de leur sélection ultérieure. En effet, dans la stratégie de Dennett, la théorie de l'évolution doit servir de théorie-relais permettant la mise en relation des états mentaux intentionnels considérés comme des *abstracta* ayant un rôle purement prédictif et des états fonctionnels des organismes, qui sont non-intentionnels mais en revanche dotés de capacités causales. Dennett mise en quelque sorte sur une forme d'harmonie préétablie entre point de vue intentionnel et point de vue de la conception, harmonie préétablie qui n'est plus d'institution divine comme dans la métaphysique classique, mais est le résultat du processus d'évolution. Or la théorie de l'évolution ne peut jouer ce rôle de théorie-relais que si elle permet de conclure à l'optimalité de fait des systèmes biologiques existants. L'évolution doit donc se comporter comme le Dieu leibnizien qui amène à l'existence celui des mondes possibles qui comprend la plus grande quantité d'essence: entre toutes les options possibles elle doit choisir la plus adaptée. Si, comme le soutient Dennett, tous les facteurs répertoriés de l'évolution, à l'exception de la sélection naturelle, sont des facteurs-amont, l'optimalité factuelle est garantie. Les facteurs-amont ne font que délimiter la sphère des possibles; on peut compter sur la sélection naturelle, qui est par définition un agent optimisateur, pour choisir le meilleur des possibles. Les facteurs-amont sont à l'évolution ce que les capacités de temps et d'espace sont à la métaphysique leibnizienne: ils définissent "l'ordre possible d'existence"<sup>3</sup>. La sélection naturelle est l'analogue laïcisé de la mathématique divine: elle garantit la sélection du système optimal tout comme la mathématique divine garantit la venue à l'existence du monde par lequel la plus grande quantité d'essence se réalise. L'optimalité des systèmes existants est la conséquence nécessaire de l'action de la sélection naturelle.

Mais que se passe-t-il si certains des facteurs autres que la sélection naturelle ne sont pas réductibles à des facteurs-amont, s'ils ne se contentent pas de délimiter le champ des variations possibles, mais agissent également en aval? Si la persistance d'un système ou d'une configuration ne peut pas toujours être imputée à la sélection naturelle, si d'autres facteurs peuvent l'expliquer, on ne peut plus conclure de la persistance d'une configuration à son optimalité. Or, il semble bien que l'hypothèse selon laquelle la sélection naturelle constituerait l'unique facteur-aval ne résiste pas à l'épreuve des faits. Le biologiste Ernst Mayr, avec lequel Dennett s'estime pourtant en "bonne compagnie", affirme ainsi qu'"En réalité, il y a des perturbations stochastiques (des événements dûs au hasard) à chaque étape du processus de sélection naturelle" (Mayr, 1988: 157). Mayr se réclame bien d'un adaptationnisme modéré, mais ce n'est pas celui de Dennett. L'adaptationnisme modéré de Mayr

---

<sup>3</sup> Cf. en particulier *De l'Origine Radicale des Choses*, dans Leibniz, 1972: 340 sq.

consiste en l'introduction d'éléments de nécessité dans le processus de la variation *mais aussi* d'éléments de hasard dans le processus de sélection.

Autrement dit, il n'est pas nécessairement vrai que si un gène conduisant à un meilleur phénotype que les autres gènes existants est présent dans une population, ce gène sera automatiquement sélectionné et que le phénotype se répandra dans la population. Stich (1990) décrit une série de phénomènes qui constituent des contre-exemples à la thèse selon laquelle l'évolution est toujours strictement adaptative. Des processus tels que ceux de la dérive génétique ou de la dérive méiotique qui introduisent des changements dans la fréquence des gènes coexistent avec le processus de la sélection adaptative et peuvent conduire à l'élimination d'un gène plus adaptatif en faveur d'un gène qui l'est moins. Le phénomène de la supériorité hétérozygote peut aussi expliquer l'incapacité de la sélection naturelle à éliminer d'une population des phénotypes non optimaux (dans ce cas, la moitié de la descendance des individus hétérozygotes optimaux sera composé d'individus homozygotes, non optimaux) et la persistance d'un polymorphisme équilibré. Le cas le plus connu et le plus dramatique est celui du gène de l'anémie falciforme ou drépanocytose. Les individus homozygotes souffrent d'une forme grave d'anémie souvent fatale dès l'enfance. En revanche, les individus hétérozygotes ne souffrent pas d'anémie et sont extrêmement résistants à la malaria. Inversement, comme le montre Kitcher (1987), il est possible qu'un génotype homozygote AA donnant lieu au phénotype le plus apte reste minoritaire ou soit éliminé d'une population, si A est peu répandu ou si les génotypes hétérozygotes contenant A donnent des phénotypes peu aptes. Enfin, il peut y avoir polymorphisme tout simplement parce qu'il n'existe pas de phénotype unique qui soit meilleur que tout autre.

L'adaptationnisme modéré compatible avec les données nouvelles apparues en théorie de l'évolution préserve l'idée que la sélection du plus apte est l'un des facteurs importants de l'évolution mais fait coexister cette idée avec l'idée que des facteurs stochastiques interviennent également dans le processus de sélection. Cette forme d'adaptationnisme modéré ne garantit donc pas que les systèmes les mieux conçus seront automatiquement sélectionnés, et même si l'on écarte les perturbations stochastiques, elle ne garantit pas non plus l'existence d'une solution optimale unique. Il apparaît donc que le recours à une théorie adaptationniste de l'évolution ne permet pas de résoudre les difficultés que rencontre le réalisme modéré de Dennett. Pour réaliser son programme, Dennett misait sur un parallélisme entre rationalité et optimalité et le principe de sélection naturelle était censé apporter une double garantie à ce parallélisme. De manière générale, il devait garantir l'optimalité de conception des systèmes biologiques. Plus particulièrement, il devait garantir l'optimalité des systèmes cognitifs humains et leur rationalité avec

pour conséquence que, dans le processus d'attribution d'états intentionnels, tant le système interprétant que le système interprété se comportent rationnellement. Or, nous l'avons vu l'évolution ne garantit ni l'optimalité de conception des systèmes, ni l'unicité du choix. Qu'en résulte-t-il?

Le parallélisme postulé par Dennett entre point de vue intentionnel et point de vue de la conception n'est pas établi. Dennett ne peut donc se fonder sur ce parallélisme pour expliquer comment des états mentaux auxquels il refuse un rôle causal peuvent conserver un intérêt prédictif. Si la théorie de l'évolution n'est pas le relais attendu, il faut trouver une autre explication. L'explication qui vient immédiatement à l'esprit est l'explication réaliste. Dennett la récuse. Mais ses arguments à l'encontre des positions réalistes ont partie liée avec sa vision de l'évolution.

Nous l'avons vu, l'une des raisons de l'opposition de Dennett à la conception réaliste est son adhésion à la thèse de l'indétermination des contenus. Cette indétermination des contenus peut selon Dennett être dérivée d'une indétermination fonctionnelle sous-jacente. Or, nous avons montré que l'argument de Dennett en faveur de l'indétermination fonctionnelle repose sur un postulat d'unicité qui n'est pas biologiquement fondé. Même si l'on admet avec Dennett que les systèmes cognitifs sont le produit de l'évolution, on ne peut affirmer que la théorie de l'évolution fournit un ancrage biologique à la thèse de l'indétermination des contenus. La théorie de l'évolution pourrait bien, au contraire, miner la thèse d'indétermination des contenus et avec elle la théorie des systèmes intentionnels.

Pour pouvoir affirmer que plusieurs interprétations du comportement d'un système sont compatibles avec la totalité des faits observables, l'interprète devrait être omniscient et parfaitement rationnel. Il devrait connaître toutes les contraintes tant internes qu'externes qui s'exercent sur le comportement du système et en déduire toutes les interprétations compatibles avec ces contraintes. Mais si le système cognitif de l'interprète est le produit de l'évolution, la rationalité et les capacités de ce système cognitif seront nécessairement limitées. L'interprète ne sera ni omniscient ni totalement rationnel. Par conséquent, l'existence de plusieurs interprétations concurrentes du système interprété pourra être imputée à la méconnaissance par l'interprète de certaines des contraintes agissantes ou, si l'on fait l'hypothèse que la totalité des faits sont connus, à des limitations de sa rationalité. En outre, si la nature favorise la diversité génétique et si les systèmes cognitifs sont dans une large mesure sous contrôle génétique, il serait étonnant que nous ne rencontrions pas une grande diversité cognitive. Si l'on admet cette diversité cognitive, il est possible que l'indétermination soit dans l'oeil de l'interprète et non dans le système interprété. Si les deux systèmes en présence (interprète et interprété) sont par trop dissemblables et

si leurs modes de raisonnement sont trop différents, il est possible qu'un comportement dont l'interprète est incapable de donner une interprétation déterminée, soit en revanche parfaitement déterminé du point de vue du système interprété. Les succès de la psychologie naïve indiqueraient alors non pas, comme le soutient Dennett, que les systèmes cognitifs sont optimalement conçus, mais que, dans leur diversité, ils ne sont pas totalement disparates. Les cas d'indétermination de l'interprétation ne récuseraient pas le mythe de la réalité objective, concrète, des croyances et désirs, ils récuseraient le mythe de l'uniformité cognitive.

Contrairement, aux espoirs de Dennett, la théorie de l'évolution n'offre pas une version laïcisée, scientifiquement acceptable, de la théorie de l'harmonie préétablie. La sélection naturelle n'institue pas de coïncidence miraculeuse entre les ordres disjoints des causes et des raisons et ne nous dispense pas d'avoir à livrer une théorie de l'intentionnalité. Plutôt que de rêver aux délices de l'harmonie préétablie, il nous faut découvrir les médiations concrètes qui rendent compte du rapport significatif que les sujets entretiennent avec le monde. L'attention portée à la théorie de l'évolution et à la biologie peut nous préserver de certaines errances, nous mettre en garde contre certains préjugés déterministes, contre une croyance naïve à l'uniformité cognitive, mais nous ne pouvons nous décharger sur elle de certains des problèmes les plus difficiles de la philosophie de l'esprit.

## **BIBLIOGRAPHIE**

AMUNDSON, R., 1990, "Doctor Dennett and Doctor Pangloss: Perfection and selection in biology and psychology", *Brain and Behavioral Sciences*, 13, 557-584.

CHURCHLAND, P. S., 1986, *Neurophilosophy*, Cambridge, Mass.: MIT Press.

DENNETT, D., 1981, *Brainstorms*, Cambridge, Mass.: MIT Press.

- DENNETT, D., 1987, *The intentional Stance*, Cambridge, Mass.: MIT Press.  
Traduction française, 1990, *La stratégie de l'interprète*, Paris: Gallimard.
- DRETSKE F., 1981, *Knowledge and the Flow of Information*, Cambridge, Mass.: MIT Press.
- DRETSKE, F., 1988, *Explaining Behavior*, Cambridge, Mass.: MIT Press.
- ENDLER, J., 1983, "Natural and sexual selection on color patterns in poecillid fishes", *Environmental biology of Fishes*, 9: 173-90.
- FODOR, J. A., 1981, *Representations*, Brighton: Harvester Press.
- FODOR, J. A., 1985, "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-mecum", *Mind* 373: 76-100.
- FODOR, J. A., 1987: *Psychosemantics*, Cambridge, Mass.: MIT Press.
- GELMAN, R., SPELKE, E. et E. MECK, 1983, "What preschoolers know about animate and inanimate objects", dans *The acquisition of symbolic skills*, ed. D. Rogers et J. Sloboda, New-York: Plenum Publishing Corp.
- GOULD, S. J. et R. LEWONTIN, 1979, "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme". *Proc Roy. Soc. London* B205: 581-598.
- HIRSCHFELD, L., 1988, "On acquiring social categories: cognitive development and anthropological wisdom", *Man*, 23: 611-38.
- JACOB, F., 1977, "Evolution and Tinkering", *Science* 196: 1161-1166.
- KITCHER, P., 1985, *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge, Mass.; MIT Press.
- KITCHER, P. 1987, "Why not the best?", dans *The Latest on the Best: Essays on Optimality and Evolution*, ed. J. Dupre. Cambridge, Mass.: MIT Press.
- KITCHER, P. et Ph. KITCHER, 1988, "The devil, the details and Dr Dennett", *Behavioral and Brain Sciences*, 11, 517-518.
- LEIBNIZ, G. W., 1972, *Oeuvres*, Paris: Aubier Montaigne.
- LEWONTIN, R., C., 1983, "The organism as the object and subject of evolution", *Scientia* 118, pp. 63-82.
- MAYR, E., 1988, *Towards a new philosophy of biology*, Cambridge, Mass.: Harvard University Press.
- STICH, S. P., 1983, *From Folk Psychology to Cognitive Science: The case against belief*, Cambridge: MIT Press.
- STICH, S., P., 1990, *The Fragmentation of Reason*, Cambridge, Mass.: MIT Press.
- TURIEL, E. et P. DAVIDSON, 1986, "Heterogeneity, inconsistency and asynchrony in the development of cognitive structures", dans I. Lewin, ed., *Stage and Structure*, Norwood, NJ: Ablex.